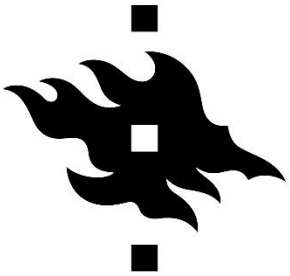
University of Helsinki
Faculty of Biological and Environmental Sciences
Master's Programme in Genetics and Molecular Biosciences



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Host range of antibiotic resistance genes in influent, effluent and dried sludge of a wastewater treatment plant

Denise Pezzutto

2019

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty	Koulutusohjelma – Utbildingsprogram – Degree Programme	
Biological and Environmental Sciences	Genetics and Molecular Biosciences	
Tekijä – Författare – Author		
Denise Pezzutto		
Työn nimi – Arbetets titel – Title		
Host range of antibiotic resistance genes in influent, effluent and dried sludge of a wastewater treatment plant		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track		
Genetics and Genomics		
Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's thesis	12/2019	55

Tiivistelmä – Referat – Abstract

Antimicrobial resistance is an emerging concern at the global scale, threatening the effectiveness of antibiotics in treating bacterial infections. Among anthropogenically impacted environments, wastewater treatment plants have been indicated as possible reservoirs of antibiotic resistance genes, putative hotspots for their horizontal gene transfer, and a source of their dissemination to the environment. Generally, the abundance of antibiotic resistance genes is reduced during the wastewater treatment process. However, some genes were shown to be enriched in purified effluent water and dried sludge, which are then released to the environment, compared to influent water. Also, the taxonomy of the hosts carrying antibiotic resistance genes could change as a result of horizontal gene transfer events.

The aim of this study was to analyse and compare the host range of a series of antibiotic resistance genes in influent water, effluent water and dried sludge collected from the Viikinmäki wastewater treatment plant in Helsinki, Finland, by applying Emulsion, Paired Isolation and Concatenation PCR (epicPCR). EpicPCR is a method that can link a gene of interest to the 16S rRNA gene from the genome of the host bacterium, without any cultivation step. The abundance of the hosts was also evaluated by sequencing the 16S rRNA gene from the whole bacterial community.

In several cases, the target antibiotic resistance genes (*bla*$_{IMP}$, *bla*$_{NDM}$, *ermB, ermF, sul1* and *strB*) were carried in effluent water and dried sludge by taxa that were not hosting them in influent water, suggesting that horizontal gene transfer events might have occurred during the treatment. All the examined genes were detected both in abundant and in rare taxa, including genera that also comprise pathogenic species, such as *Arcobacter* and *Acinetobacter*. Some of the detected hosts were not previously known to show resistant phenotypes, namely members of the family Methylophilaceae.

These results corroborate the idea that wastewater treatment plants might be hotspots for the horizontal gene transfer of resistance determinants, and potentially disseminate antibiotic resistant pathogens to the environment. However, in order to ensure the accuracy of the results, the limits of epicPCR as a method need to be identified and addressed.

Avainsanat – Nyckelord – Keywords

Antimicrobial resistance, sewage, horizontal gene transfer, microbial ecology, epicPCR, 16S rRNA gene sequencing

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors

Antti Karkman, Marko Virta

Säilytyspaikka – Förvaringställe – Where deposited

E-thesis

Muita tietoja – Övriga uppgifter – Additional information

TABLE OF CONTENTS

ABBREVIATIONS

AMR	Antimicrobial Resistance
ARG	Antibiotic Resistance Gene
ARB	Antibiotic Resistant Bacteria
ASV	Amplicon Sequence Variant
epicPCR	Emulsion, Paired Isolation and Concatenation PCR
ESBL	Extended-spectrum β-lactamase
HGT	Horizontal Gene Transfer
MGE	Mobile Genetic Element
MIC	Minimum Inhibitory Concentration
MLSB	Macrolide, Lincosamide and Streptogramin B
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
WWTP	Wastewater Treatment Plant

INTRODUCTION

THE ANTIBIOTIC RESISTANCE CRISIS

Antibiotics are antimicrobial compounds that existed for tens of thousands of years in the environment, produced by bacteria and fungi to inhibit the growth of their commensals. Parallelly, antibiotic resistance evolved as a defence mechanism, and therefore it existed way before the clinical use of antibiotics carried out by humans (Pal et al., 2016). Indeed, several genes that confer resistance to antibiotics have been found in ancient samples and in modern environments that were never in direct contact with anthropogenic sources, showing the ubiquity and ancientness of resistance genes (Crofts et al., 2017). The human use of antibiotics, which began in the 1930s, brought a revolution in the medical field. Most infections that were previously deadly became very easy to treat, and this has been the case for decades. Already in the years after the first antibiotics were commercialized, it was observed that bacteria could develop resistance to them. This was initially not considered a major issue, even when it was reported that resistance could be transmitted to other strains, and therefore be spread (Davies, 2007). Initially, as a temporary solution to the emergence of antibiotic resistance, research efforts focused towards the development and commercialization of new antibiotics (Crofts et al., 2017). Even after the threat posed by the emerging and spread of antibiotic resistance was recognized, the use of antibiotics kept constantly growing.

Since then, the frequency of antibiotic resistance genes (ARGs) and antibiotic resistant bacteria (ARB) has continued to dramatically increase, posing threats to human and animal health. Indeed, together with the frequency of ARGs, the frequency of antibiotic resistant bacterial infections increases. As antibiotics are not an effective therapeutic strategy against them, these infections are more difficult, and sometimes even impossible, to treat (Crofts et al., 2017). In the absence of concrete actions to tackle this problem, a post-antibiotic era will occur, meaning that infective diseases that are now commonly treated with antibiotics will become deadly again (O'Neill, 2016). It was estimated that, already nowadays, at least 700 000 people worldwide die every year due to antibiotic resistant infections (O'Neill, 2014). This figure is projected to rise to 10 million annual deaths in 2050, in the absence of adequate countermeasures (O'Neill, 2016).

Nowadays, when looking at the global scale, the spread of already existing ARGs seems to have a larger impact on the total increase in frequency of AMR compared to their selection or the emergence of new resistance determinants (Collignon et al., 2018). Current antibiotic consumption has not been found to correlate with antimicrobial resistance levels (Collignon et al., 2018). On the other hand, the quality of infrastructure and governance strongly correlates with antimicrobial resistance levels, showing that, for instance, better sanitation and access to clean water correlate with lower

antimicrobial resistance (Collignon et al., 2018). This is to indicate that reducing antibiotic consumption will not be enough to address the current antibiotic resistance crisis (Collignon et al., 2018), and it further underlines the impact of antibiotic resistance transmission routes on the issue.

THE ENVIRONMENTAL RESISTOME

Antibiotic resistance is often regarded as a clinical issue, since its effects are mostly visible when resistant bacteria cause infections in humans or animals. However, antibiotic resistant bacteria are consistently found in the human, animal and environmental microbiome, which act as reservoirs for ARGs (Pal et al., 2016). The overall set of ARGs that is present in an environment constitutes its resistome (Crofts et al., 2017), and its richness and composition are important factors to consider when assessing the risks that it might pose to human health (Pal et al., 2016). Pathogens are capable of acquiring ARGs from other bacteria carrying them, through the spread of mobile genetic elements (MGEs) (Pal et al., 2016). For this reason, the presence of ARGs in the environment is not to be overlooked.

The risk connected with an environment as a possible source of dissemination of ARGs to human pathogens depends on the abundance of ARGs in it, but also on the diversity both of the genes themselves and of the microbial community (Pal et al., 2016). Specifically, the environmental microbiome has a greater abundance and diversity of ARGs compared to human and animal microbiomes, also due to the higher level of diversity in the microbial community (Pal et al., 2016). This further suggests that the environment, besides being a reservoir of ARGs, is a possible source of dissemination of clinically relevant ARGs to pathogenic bacteria.

The ARG content and host range of each specific environment should be investigated, to evaluate its contribution to the AMR problem. In this perspective, human impacted environments are of particular interest because they bring together ARB that originate from anthropogenic sources with those that are present in the environment.

WASTEWATER TREATMENT PLANTS AS HOTSPOTS FOR AMR

Among human impacted environments, urban wastewater treatment plants (WWTPs) are especially interesting in the study of AMR, because they receive sewage of human origin and ultimately release it into the environment (Karkman et al., 2018).

WWTPs receive sewage waters coming from different urban sources, including but not limited to households, hospitals and industries. Since a great part of the antibiotics used by humans ends up in sewage waters, antibiotics are present inside WWTPs, even though in low concentration. Together with antibiotics, ARB enter WWTPs through influent sewage waters. Therefore, inside WWTPs, bacteria from different sources are found in proximity to each other and might be able to exchange

genetic material, including ARGs, with horizontal gene transfer (HGT) events (Karkman et al., 2018). In addition, the presence of antibiotics and other compounds in wastewaters might exert a selective pressure on the bacterial community (Karkman et al., 2018). For these reasons, WWTPs have been indicated as possible hotspots for the selection of ARGs and their transmission between different bacterial species (Rizzo et al., 2013).

The treatment process leads to the formation of by-products, namely purified effluent water and excess sludge. The purified water is generally released to natural water bodies, while the excess sludge, after being dried, is usually land applied. Because of these by-products, WWTPs are also a source of dissemination of ARGs to the environment (Rizzo et al., 2013).

Sewage waters are also a remarkable sample type to perform surveillance of antimicrobial resistance, because they are relatively easy to obtain and no ethical permission is required (Hendriksen et al., 2019). With this in mind, WWTPs could be used for monitoring the levels of antimicrobial resistance in a given area.

MICROBIAL COMMUNITY OF WWTPS

As previously mentioned, the diversity of the microbial community in an environment affects the risk associated with it as a reservoir of ARGs (Pal et al., 2016). For this reason, possible changes in the bacterial community of wastewaters need to be considered.

The microbial community characterizing WWTPs consists of bacteria that take part in the treatment process itself. Activated sludge contains a mixture of different prokaryotes which carry out distinct functions, and it is a widely used tool in wastewater treatment for removing nutrients such as carbon, nitrogen and phosphorus, as well as micropollutants (Wu et al., 2019). Worldwide, the number of different bacterial species included in activated sludge has been estimated to be ~10^9, of which only 10^4 are known (Wu et al., 2019). Globally, *Dokdonella, Zoogloea, Nitrospira, Haliangium, Rhodoferax, Dechloromonas, Arcobacter, Cloacibacterium, Turneriella, Zymomonas*, Candidatus *Accumulimonas, Sulfuritalea, Acinetobacter* and Candidatus *Accumulibacter* have been reported to be the core genera in activated sludge (Wu et al., 2019). Nonetheless, the core operational taxonomic units (OTUs) that are shared between WWTPs worldwide only account for about 0.5% of the total number of OTUs (Wu et al., 2019), suggesting a large degree of variation in activated sludge composition of different WWTPs.

In addition to the resident community, different bacteria are also entering WWTPs with influent water. Sewage waters are collected from different urban sources, and biofilms growing inside the sewage transport system can also be carried downstream the sewage pipes (Guo et al., 2019; Mclellan et al., 2015). Overall, human fecal bacteria only account for approximately 7-15% of the OTUs found

in sewage water communities, while over 20% are distinctive of soil microbiota (Guo et al., 2019). In studies conducted in the USA, the most abundant bacterial genera in WWTPs were reported to be *Acinetobacter, Aeromonas, Arcobacter, Pseudomonas* and *Trichococcus* (Mclellan et al., 2015). Analogously, WWTPs located in the Helsinki area receive sewage waters in which Campylobacteriales, Clostridia, Bacteroidales and Pseudomonadales are the main bacterial orders (Hultman et al., 2018). Finally, when investigating the influent community of WWTPs, the fact that it is subject to both temporal and geographic variation needs to also be considered (Guo et al., 2019). In effluent waters, the bacterial community composition depends on the treatment process of the specific WWTP. As an example, it was shown that the effluent communities of two WWTPs in the Helsinki area were significantly different, even though their influent waters did not exhibit a statistically significant difference in their composition (Hultman et al., 2018).

SELECTION OF RESISTANCE GENES IN WWTPS

It has been presumed that antibiotics that are present in sewage waters could exert a selection pressure leading to the enrichment of the resistant phenotypes during the treatment (Figure 1). In WWTPs, antibiotics are present at sub-MIC (minimum inhibitory concentration) levels, meaning that their concentration is not high enough to be lethal for susceptible bacteria (Sandegren, 2014). However, even at sub-MIC concentrations, the presence of antibiotics can give a competitive advantage to resistant bacteria, promoting their selection in the community despite not directly killing the susceptible ones (Sandegren, 2014). Low concentrations of antibiotics have been shown to select for resistant bacteria in simple communities, but their effects on complex communities like the ones that characterize WWTPs have not yet been extensively studied (Karkman et al., 2018).

Overall, the role of selection in enriching certain ARGs is not yet clear. Nonetheless, it is likely to vary based on the specific treatment plant, on the microbial community, on the chemical composition of the sewage water, and ultimately depends on the specific ARG.

HORIZONTAL GENE TRANSFER OF ARGS IN WASTEWATERS

While selection can only increase the frequency of resistant bacteria once they already carry a given ARG, one way for an AMR determinant to arise in a new host is horizontal gene transfer.

Horizontal gene transfer (HGT) is the exchange of genetic material between two organisms without reproduction. In bacteria, the mechanisms by which HGT can occur are transformation, transduction and conjugation (Thomas and Nielsen, 2005) (Figure 1).

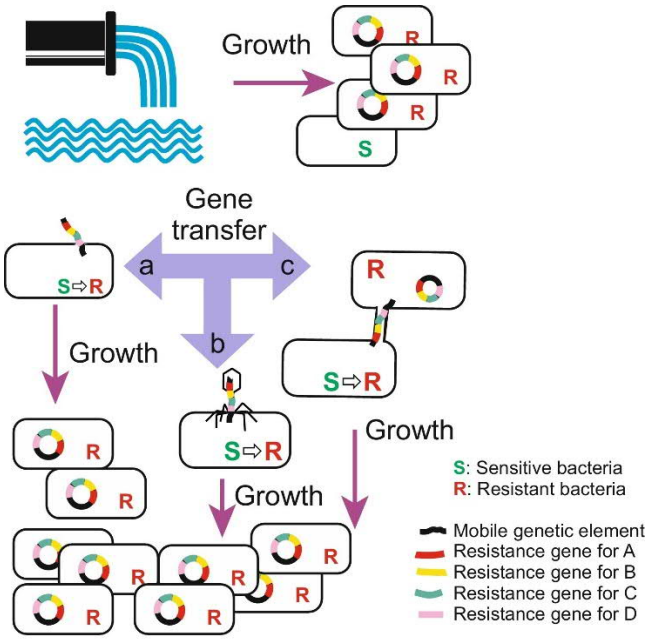


Figure 1 Representation of selection and horizontal gene transfer mechanisms in wastewaters. Growth: resistant bacteria (R) can be selected in the presence of antibiotics and other compounds, outgrowing sensitive bacteria (S). Different resistance genes can be co-selected. Gene transfer: antibiotic resistance genes may be acquired by sensitive bacteria by transformation (a), transduction (b) or conjugation (c). This promotes the spread of antibiotic resistance genes in the bacterial community. Figure from Karkman et al., 2018. (Licence obtained via Copyright Clearance Center's RightsLink® service. ®Elsevier Ltd.)

In transformation, the cell intakes free DNA from the surrounding environment. In transduction, DNA is introduced in the cell with a bacteriophage infection. Lastly, conjugation involves a donor cell transferring DNA to a recipient cell, physically contacting it through a pilus (Thomas and Nielsen, 2005).

ARGs are often contained in mobile genetic elements (MGEs), such as plasmids and transposons (Partridge et al., 2018). In addition, MGEs can include other genetic elements such as integrons, that are known to commonly contain gene cassettes with even multiple ARGs (Partridge et al., 2018). The main characteristic feature of MGEs is that they are easily mobilized intracellularly and intercellularly (Partridge et al., 2018). Due to this property, the localization of ARGs on MGEs highly increases their chances of being transferred between bacteria, thus spreading both the resistant phenotypes and their genetic determinants.

WWTPs are relevant for the spread of ARGs because they bring different bacteria near each other, increasing the chances for HGT to occur. However, in addition to this, MGEs are also mobilized more effectively in the presence of antibiotics, which are found in wastewaters. For instance, HGT of the transposon Tn*916* is known to be induced by tetracycline (Showsh and Andrews, 1992). Sub-MIC concentrations of antibiotics, like the ones that are found in wastewaters, have been shown to promote HGT in small communities (Cairns et al., 2018). However, this does not necessarily apply to complex

environmental communities, so more research is needed to shed light on the role of HGT in WWTPs. One way to identify putative HGT events would consist in investigating the host range of ARGs before and after the process, and then comparing the results.

RELEASE OF ARGS FROM WWTPS INTO THE ENVIRONMENT

Overall, WWTPs are effective in removing ARGs from sewage waters prior to release into the environment (Karkman et al., 2018). The abatement of ARGs abundance in effluent waters compared to influent waters has been shown to mostly depend on the reduction of the overall bacterial load (Laht et al., 2014).

However, even though their abundance decreases with the treatment process, ARGs can still be detected in effluent waters and sludge, and this poses a risk of ARGs pollution to the environment (Karkman et al., 2018). For instance, effluent wastewater discharges have been shown to increase the abundance of ARGs in the downstream water environment (Cacace et al., 2019). The increase in ARGs abundance in environments receiving wastewater discharges has been imputed to fecal pollution, meaning that the bacterial load that is remaining after the treatment is solely responsible for the spread of AMR in the environment (Karkman et al., 2019).

In addition to the overall ARGs load before and after the treatment process, changes in frequency of individual genes are also relevant. In this respect, quantitative studies targeting individual ARGs have shown that some genes are enriched during the process (Bengtsson-Palme et al., 2016; Karkman et al., 2016). It is currently not known whether this enrichment of certain genes is the result of selection for the resistant phenotypes, or if their reported increase in frequency is a consequence of the increase in frequency of their hosts, which can be due to other selective pressures in the WWTP. HGT might also lead to an enrichment of ARGs, by broadening their host range. In order to understand the causes underlying the increase in frequency of selected genes, investigating their host range would be a necessary step.

SPECIFIC ARGS IN WWTPS

Sewage waters contain a broad range of ARGs (Hendriksen et al., 2019). Worldwide, the most abundant ones are the genes that confer resistance to macrolides-lincosamides-streptogramin B (MLSB) (e.g. *ermF*), tetracyclines (e.g *tetM*), aminoglycosides (e.g. *aadA, strB*), β-lactams (e.g. bla_{OXA}), and sulphonamides (e.g. *sul1*) (Hendriksen et al., 2019). Similar results were obtained when looking at the ARGs load of WWTPs across the European continent, with genes associated with aminoglycoside-, β-lactam-, MLSB-, tetracycline- and multidrug- resistance being found in all influent samples (Pärnänen et al., 2019). Globally, the most abundant ARGs were found to be *aadA,*

bla$_{OXA}$, erm(B), erm(F), mef(A), mph(E), msr(D), msr(E), strA, strB, sul1/sul3, tet(39), tet(Q), and *tet(W)* (Hendriksen et al., 2019).

Due to their abundance and widespread presence, some genes were also suggested as possible markers of AMR abundance in wastewaters: *aadA, cmxA, bla$_{OXA}$, qacEΔ1, ermF, qnrSrtF11, sul1, tetQ, ISPps, int1* and *tnpA* (Pärnänen et al., 2019). The class I integron *int1* and the transposase gene *tnpA* are considered important markers of ARGs pollution because often associated with ARGs. Also, they are markers for the mobility potential of ARGs in a given environment. Specifically, the *int1* gene is linked to ARGs such as *sul1* and *qacEΔ1* (Gillings et al., 2008). Both *sul1* and *qacEΔ1*, together with the insertion sequence ISSm2, were present both in influent and in effluent waters in all samples collected from European WWTPs (Pärnänen et al., 2019). Similarly, almost all transposases were shown to be highly abundant in sewage waters in the Helsinki WWTP, and one variant of the transposase gene *tnpA* (*tnpA-04*) was enriched in effluent water and dried sludge compared to influent water (Karkman et al., 2016).

Although not so abundant, genes that have been indicated as clinically relevant and/or that have emerged in recent times are also found in sewage waters. For instance, *bla$_{CTX-M}$, bla$_{NDM}$, mcr* and *optrA* have been detected in WWTPs across the world (Hendriksen et al., 2019), and *bla$_{IMP}$* and *vanA* have been found in some influent samples in different European countries (Pärnänen et al., 2019).

CHOICE OF ARGS FOR THE PRESENT STUDY

For this study, a list of eight genes was chosen for their host range to be investigated. At first, twelve genes were shortlisted, and afterwards a subset of eight was designed based on their presence in the specific samples, the efficacy of their primers, and their biological relevance. As the samples to be used were collected from Viikinmäki WWTP in Helsinki, Finland, a study that quantified ARGs in the same WWTP was used as the main source of information for choosing the target genes (Karkman et al., 2016). For a more comprehensive picture, genes conferring resistance to several antibiotic classes were chosen.

For the macrolide-lincosamide-streptogramin B (MLSB) antibiotic class, *ermB* and *ermF* genes were selected. The first, *ermB*, was reported to be one of the most abundant ARGs in wastewaters worldwide (Hendriksen et al., 2019), and it was also detected in Viikinmäki WWTP (Karkman et al., 2016). Its known host range includes several clinically relevant genera, among which *Acinetobacter, Bacillus, Escherichia, Klebsiella, Neisseria, Enterobacter* and *Pseudomonas* (Roberts et al., 1999; Roberts, 2008). *ErmF* was also indicated as one of the most abundant ARGs in wastewaters (Hendriksen et al., 2019). In addition, it was proposed as a possible marker for AMR in sewage waters (Pärnänen et al., 2019). In Viikinmäki WWTP, *ermF* was enriched in dried sludge, exhibiting a 25-

fold increase in its frequency (Karkman et al., 2016), which makes it an interesting target gene. Genera hosting *ermF* were reported to be, among others, *Bacteroides, Neisseria, Prevotella, Ruminococcus, Shigella* and *Staphylococcus* (Roberts et al., 1999; Roberts, 2008).

Two sulphonamide-resistance genes were targeted, *sul1* and *sul2*. *Sul1* has been indicated as a marker of AMR frequency (Pärnänen et al., 2019), and it is one of the most abundant ARGs in sewage waters (Hendriksen et al., 2019). Also, the gene was already detected in the WWTP under examination (Karkman et al., 2016; Laht et al., 2014). It is also liked to the class I integron gene *int1* (Gillings et al., 2008), which can promote its mobility. *Sul2* was also detected in Viikinmäki WWTP (Karkman et al., 2016; Laht et al., 2014), where it was found to be enriched in effluent waters compared to influent waters (Karkman et al., 2016). The increase in frequency of *sul2* during the treatment process makes it intriguing as a target gene for this study. Both *sul* genes have been detected in a vast range of genera since their discovery, among which *Escherichia, Salmonella, Pseudomonas, Acinetobacter, Aeromonas* and *Bacillus* (Phuong Hoa et al., 2008).

Given the clinical relevance of β-lactam antibiotics, three genes coding for β-lactamases were considered: bla_{IMP}, bla_{NDM} and $bla_{CTX\text{-}M}$. All these genes have been sporadically detected in sewage waters across the world (Hendriksen et al., 2019; Pärnänen et al., 2019), but not in Viikinmäki WWTP specifically.

The metallo-β-lactamase gene bla_{IMP}, conferring resistance to carbapenem and other broad spectrum β-lactams, is often located in class I integrons together with other ARGs, possibly resulting in multi-resistance (Zhao and Hu, 2011). Different gene variants have been detected in different hosts, but overall its host genera include, among others, *Pseudomonas, Acinetobacter, Citrobacter, Escherichia, Enterobacter, Klebsiella* and *Aeromonas* (Zhao and Hu, 2011).

Analogously to the IMP-type ones, NDM enzymes are also class B metallo-β-lactamases. The first documented infection by bla_{NDM}-carrying bacteria took place in 2008 in a patient that was hospitalized in New Delhi, hence the name bla_{NDM} (New Delhi metallo-β-lactamase) (Johnson and Woodford, 2013). The gene is considered to have epidemiologically originated in the Indian subcontinent, but it has rapidly spread across the globe, to the point of being detected in all continents except South America and Antarctica (Bonomo, 2017). At first detected only in *Escherichia coli* and *Klebsiella pneumoniae*, the known host range of bla_{NDM} includes now also other Enterobacteriaceae, *Acinetobacter* and *Pseudomonas*, and its fast dissemination seems to depend on conjugative plasmid transfer (Johnson and Woodford, 2013).

CTX-M enzymes are class A extended-spectrum β-lactamases (ESBLs). ESBLs confer resistance to a broad range of clinically relevant antibiotics, namely penicillins and some broad-spectrum cephalosporins (cefotaxime, ceftriaxone and ceftazidime) (Zhao and Hu, 2013). Notably, ESBLs

coded by *bla$_{CTX-M}$* genes are frequent in pathogens such as *Escherichia coli* and *Klebsiella pneumoniae* (Zhao and Hu, 2013). Other relevant *bla$_{CTX-M}$*-carrying genera are *Acinetobacter, Aeromonas, Enterobacter, Pseudomonas, Salmonella, Shigella* and *Stenotrophomonas* (Zhao and Hu, 2013).

The study included two tetracycline-resistance genes, *tetM* and *tetG*. Both were detected in sewage waters worldwide (Hendriksen et al., 2019). The list of genera harbouring *tetG* comprises *Escherichia, Pseudomonas, Salmonella, Vibrio* and *Providencia* (Chopra and Roberts, 2001; Roberts, 2005), while *tetM* is found in a broader range of genera including *Acinetobacter, Bacteroides, Enterobacter, Escherichia, Klebsiella, Neisseria* and *Pseudomonas* (Chopra and Roberts, 2001; Roberts, 2005).

The aminoglycoside-resistance gene *strB* was included in the analysis because it was reported to be one of the most frequent genes in wastewaters (Hendriksen et al., 2019). This gene is also often found in MGEs, which causes it to be easily spread (Sundin and Bender, 1996). In addition, *strB* tends to be located in a gene cluster together with *strA* and the previously mentioned *sul2* (Sundin and Bender, 1996). Among the genera in which *strB* has been detected are *Escherichia, Enterobacter, Klebsiella, Neisseria, Salmonella, Shigella* and *Pseudomonas* (Sundin and Bender, 1996).

Finally, two MGEs were targeted. As previously mentioned, the class I integron *int1* gene is relevant to the study of AMR because of its frequent association with ARGs (Gillings et al., 2008) and its ubiquity in wastewaters (Pärnänen et al., 2019). Its host range in WWTPs had already been studied (Hultman et al., 2018). The transposase-coding gene *tnpA-04* was also suggested as a marker for ARG pollution (Pärnänen et al., 2019). Interestingly, it was also found to be enriched in effluent water and dried sludge in Viikinmäki WWTP (Karkman et al., 2016).

METHODS FOR SURVEILLANCE OF AMR IN THE ENVIRONMENT

Methods that can be used to monitor the presence of ARGs can be divided in culture-based methods and molecular methods. Culture-based methods, such as those that rely on disc-diffusion or microdilution, are well established and standardized, but laborious and time consuming (Rizzo et al., 2013). While they are still valid tools for assessing the occurrence of AMR on pathogens, these methods were developed specifically for clinical settings and they are not so effective in evaluating AMR in the environment. Clinical breakpoints and cut-offs are simply not applicable to environmental samples, including wastewaters (Karkman et al., 2018). Culture-based methods are also biased towards cultivable species, and they fail to detect those bacteria that are difficult to be grown in laboratory conditions (Rizzo et al., 2013). Molecular methods, on the other hand, allow the detection of ARGs also in non-cultivable bacteria. This is particularly important when dealing with

environmental samples, as many bacterial species collected from the environment cannot be cultured (Rizzo et al., 2013).

Among molecular methods, PCR can detect the presence of ARGs while qPCR can be used to quantify their abundance in environmental samples. A drawback of PCR-based methods is that they require the design of gene-specific primers, which is an obstacle to the detection of unknown genes (Karkman et al., 2018). Metagenomics, on the other hand, does not require any previous knowledge of the resistance genes that are present in a community. By sequencing the DNA of the whole community, it can record its entire resistome, when the sequencing is sufficiently deep (Karkman et al., 2018). However, gene annotation is usually based on public sequence databases, therefore unknown genes, as well as known but not experimentally verified genes, cannot be identified (Karkman et al., 2018). Furthermore, the short reads produced by the most used sequencing platforms are not enough for linking an ARG to its host or reveal its genetic contexts (Karkman et al., 2018).

EPICPCR

The so far described methods all contribute to increasing our understanding of AMR in the environment. However, they fail to provide information on the host bacteria that are carrying the ARGs, which is crucial for understanding the dynamics of AMR dissemination. This knowledge gap can be filled by a recently developed method called Emulsion, Paired Isolation and Concatenation PCR, or epicPCR (Spencer et al., 2016).

EpicPCR is a PCR-based method that can link two genes of interest originating from uncultured single cells into one amplicon. When one of the genes is a phylogenetic marker, for instance the 16S rRNA gene, epicPCR provides information on which members of a microbial community contain the other gene. Therefore, by applying epicPCR to link the 16S rRNA gene with an ARG of interest, the host range of that ARG can be resolved.

The workflow for epicPCR is displayed in Figure 2. At first, cells collected from an environmental sample are isolated by being encapsulated in polyacrylamide beads. The cells are diluted to maximize the probability that no more than one cell is encapsulated in one bead, to avoid false positives. The PCR amplification of the two genes and their linking take place in an emulsion, to ensure that the target gene and the 16S rRNA gene are amplified from the same single cell. The product is then purified and further amplified in a nested PCR. To prevent the amplification of unfused products, blocking primers are also added to the nested PCR reaction.

EpicPCR is very promising for the study of AMR in the environment, because it does not involve any cultivation step, it has a high throughput and it can effectively link a gene to its host, which can be challenging with metagenomics.

As the method has been developed very recently, it has not been extensively used and it is still subject to optimization. In the original paper where it was presented (Spencer et al., 2016), epicPCR was used to determine the host range of the sulfate reductase gene *dsrB* in a freshwater lake. After that, the method was used to investigate the host range of *tetM*, *qacEΔ1* and $bla_{OXA\text{-}58}$ resistance genes and the integron *int1* in samples collected from Viikinmäki and Suomenoja wastewater treatment plants (Hultman et al., 2018). In a paper published in 2018, epicPCR was used to track the mobility of RP4 plasmid, conferring an antibiotic resistant phenotype, in an experimental bacterial community at low antibiotic concentrations (Cairns et al., 2018). More recently, epicPCR was applied to the study of the host range of the sulfate reductase gene *dsrB* in Tibetan saline lakes (Qin et al., 2019).

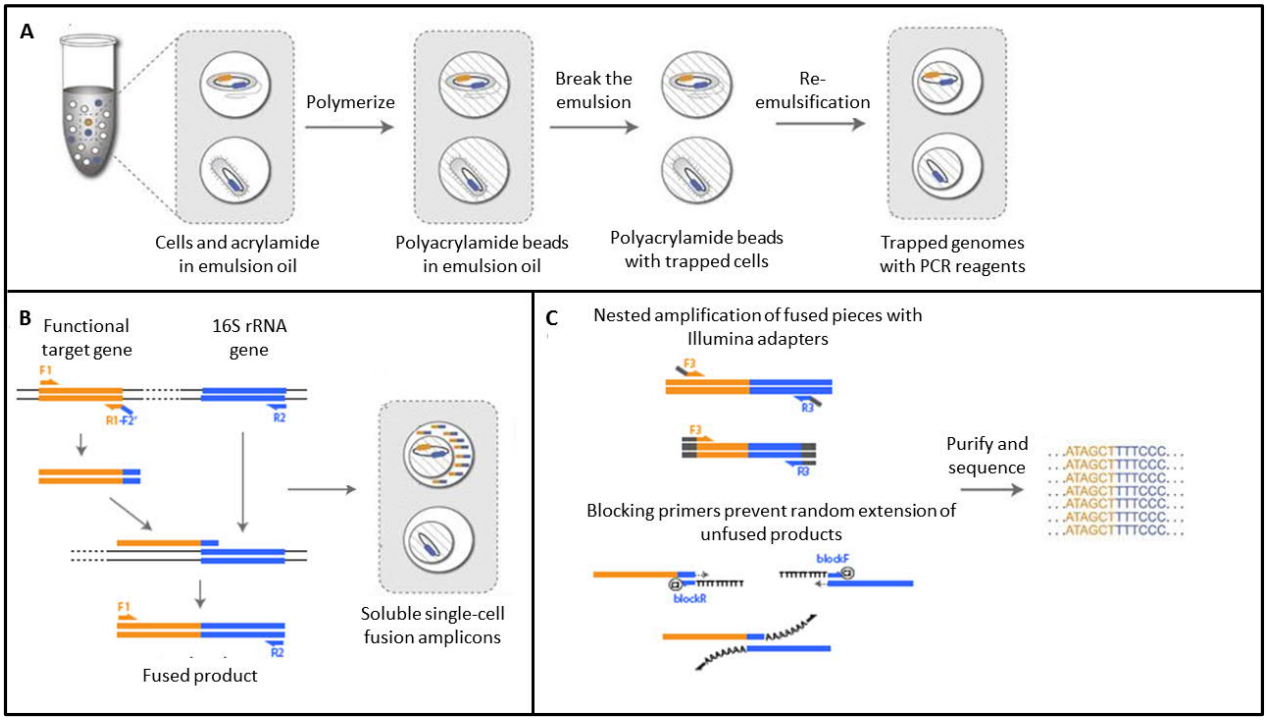


Figure 2 Workflow for Emulsion, Paired Isolation and Concatenation PCR. A) Cells are isolated by being incapsulated into polyacrylamide beads in an emulsion. After polymerization, beads are trapped in a second emulsion together with fusion PCR reagents. B) Fusion PCR leads to the amplification of the target gene fused to the 16S rRNA gene into one amplicon. c) After breaking the second emulsion, fused products are amplified with a nested PCR, which also adds the Illumina adapters to the amplicons. Blocking primers are used to prevent the extension of unfused products. Modified from Figure 1 from Spencer et al., 2016. (Open source under Creative Commons Attribution 4.0 International License, https://creativecommons.org/licenses/by/4.0/)

AIMS

This project aims at resolving the host range of eight antibiotic resistance genes in samples collected from Viikinmäki wastewater treatment plant. The target genes were chosen based on their clinical relevance or their abundance in wastewater samples, as reported in previous literature. By using epicPCR to investigate which bacteria carry each gene in influent water, effluent water and dried sludge, and comparing the results, it could be possible to track the fate of the ARGs once they exit the plant, and gather information on the possibility of ARG transmission between different hosts during the treatment process. Indeed, if a bacterial group was found to carry a gene in effluent water or dried sludge, but not in influent water, that would imply the possibility of horizontal gene transfer events occurring inside the WWTP. This would further corroborate the idea that WWTPs are hotspots for the transmission of ARGs between different bacteria. Since detected changes in the host range of an ARG might not only be explained by horizontal gene transfer, but also by variations in the relative abundance of the hosts during the treatment, the study also includes the analysis of the microbial community composition in each sample.

MATERIALS AND METHODS

SAMPLING

The samples used in this project were influent water, effluent water and dried sludge collected by the Viikinmäki wastewater treatment plant personnel on three different days (18/03/2019, 20/03/2019 and 26/03/2019). The samples were transported to the lab within two hours after collection and then immediately processed.

Viikinmäki is the largest WWTP in Finland and in the Nordic countries. In use since 1994, it collects sewage waters both from industries and from about 0.8 million residents of Helsinki and its surrounding municipalities. Approximately 85% of the incoming water is of domestic origin, while 15% derives from industries. Altogether, the WWTP treats about 270 000 m^3 of water every day, and 100 million m^3 annually. The treatment process includes a primary treatment that mechanically removes larger particles such as sand and fat, a secondary treatment involving activated sludge, and a tertiary treatment that includes biofilters for nitrogen removal (Karkman et al., 2016; Laht et al., 2014). The release site for effluent water is located in the Baltic sea, approximately 8 km southern than Helsinki. The activated sludge, on the other hand, goes through an anaerobic digestion step and is then dried and processed into soil products.

DNA EXTRACTIONS

All DNA extractions were performed in triplicate. For DNA extraction from wastewaters, 20 mL of influent water was filtered with a 0.2 µm polycarbonate filter, and 200 mL of effluent water was filtered with a 0.2 µm polycarbonate filter and a 0.45 µm cellulose filter on top of it. DNA was extracted with DNeasy PowerWater Kit (QIAGEN) after placing the filters in PowerWater Bead tubes with sterile tweezers. A negative control for each DNA extraction from influent water was prepared by placing one sterile 0.2 µm polycarbonate filter in a PowerWater Bead tube (QIAGEN). A negative control for each DNA extraction from effluent water was prepared analogously, but with an additional sterile 0.45 µm cellulose filter being placed in a PowerWater Bead tube (QIAGEN). For DNA extraction from dried sludge, DNeasy PowerSoil Kit (QIAGEN) was used with 0.25 g of dried sludge. All extractions were performed in triplicate and for each of them a negative control was prepared by using a PowerSoil tube (QIAGEN) without adding any sludge to it. The DNA extraction protocols were followed identically for samples and controls. Extracted DNA was stored at -20°C.

CELL EXTRACTIONS

All cell extractions were performed in duplicate. Cells were collected from 200 mL of effluent water by centrifuging for 40 minutes at 8 000 x *g* at 4°C. Similarly, 2 mL of influent water was centrifuged for 5 minutes at 11 000 × *g*. After removing the supernatant, the cell pellet was resuspended in 1.5 mL of 20% sterile-filtered glycerol, flash-frozen and stored at -80°C.

Dried sludge was weighted in duplicate in a 2 mL Eppendorf tube to an amount of 0.15 g. The sludge was resuspended in 700 µL of 1X PBS with a sterile wooden stick and poured to a round-bottom glass tube that was compatible with the S220 Focused-ultrasonicator (Covaris). The tubes were vortexed, placed to the sonication machine and sonicated with the following settings: duty cycle 1%, intensity 0.1, cycles per burst 100. After the treatment, 1 mL of Histodenz™ solution (Sigma-Aldrich) was added to a 2 mL Eppendorf tube, the samples were vortexed and about half of the sample was poured on top of the Histodenz™ solution. The tubes were centrifuged for 20 minutes at 5 000 x *g* at 4°C. The top and middle phases, containing bacterial cells but no larger particles, were transferred to a new tube. The same procedure was repeated with the remaining part of the sample. After the transferred sample was centrifuged for 10 minutes at 13 000 x *g* at 4°C, the supernatant was removed and the pellet resuspended in 1 mL of 20% sterile-filtered glycerol, flash-frozen and stored at -80°C.

16S rRNA GENE SEQUENCING

The V3-V4 region of the 16S rRNA gene was amplified for sequencing. The templates were the three DNA extraction replicates for all samples and the respective DNA extraction negative controls. The PCR setup included one positive control and two negative controls, which were pooled after the reaction. Reactions that did not result in a visible product when visualized with E-Gel™ EX Agarose Gels, 2% (Invitrogen) were repeated a second time, with two negative controls that were combined afterwards and one positive control.

The primers used were 341F and 785R with Illumina TrueSeq adapter sequences at the 5´end. The forward primer solution was an equimolar mix of primers 341F-1, 341F-2, 341F-3 and 341F-4 (Appendix 1), differing by a few base additions to the 5' end of the primer. Similarly, the reverse primer was an equimolar mix of primers 785R-1, 785R-2, 785R-3 and 785R-4 (Appendix 1). The concentration of the primer mixes in each PCR reaction was 0.2 µM. The other reagents were dNTPs 0.2 µM (BioNordika), GC buffer 1X (Thermo Fisher Scientific), DMSO 25% (Thermo Fisher Scientific), Phusion polymerase 0.02 U/µL (Thermo Fisher Scientific). Each 25 µL reaction contained an amount of template DNA ranging from 40 ng to 60 ng, except in the case of the DNA extraction negative controls. The PCR cycling conditions were the following: initial denaturation at 98°C for 30 seconds, 15 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 30 seconds and extension at 72°C for 10 seconds, and a final extension step at 72°C for 5 minutes.

The PCR products were sequenced with Illumina's MiSeq NGS system at the Institute of Biotechnology of University of Helsinki.

16S rRNA GENE ANALYSIS

Adapters and primers were removed from the 16S rRNA gene reads by using cutadapt (Table 1B), and quality control was performed with FastQC (Table 1B). The DADA2 package (Table 1A) was then used to obtain an amplicon sequence variant (ASV) table in R (Table 1A). Based on the quality control results, forward reads were truncated at position 300 and reverse reads at position 230 by using the filterAndTrim command. Simultaneously, reads were filtered with a maximum number of expected errors equal to 6. Forward and reverse reads were merged with the mergePairs command. The ASV table was constructed from the merged reads with the makeSequenceTable command. Chimeras were removed with the removeBimeraDenovo command. Taxonomic classification was carried out with the assignTaxonomy command, using the Silva version 132 database as the reference (Quast et al., 2013).

Table 1. Software used in this study

A) R and its packages

Software/package	Version	Source
R	3.6.1	R Core Team, 2017
DADA2	1.12.1	Callahan et al., 2016
phyloseq	1.28.0	McMurdie and Holmes, 2013
VEGAN	2.5.5	Dixon, 2003
ggplot2	3.2.1	Wickham, 2016

B) Command line tools

Software/command line tool	Version	Source
Cutadapt	1.10	Martin, 2011
FastQC	0.11.8	Andrews, 2011
mothur	1.40.5	Schloss et al., 2009
PEAR	0.9.6	Zhang et al., 2014
VSEARCH	2.6.0	Rognes et al., 2016
BLAST	2.6.0	Altschul et al., 1990
MUSCLE	3.8.31	Edgar, 2004
FastTree	2.1.9	Price et al., 2010
Anvi'o	5.5.0	Eren et al., 2015

Taxonomic data were processed with the phyloseq package (Table 1A) in R. Alpha diversity was calculated with Shannon index with phyloseq. Beta diversity was calculated with a permutational multivariate analysis of variance, using Bray-Curtis dissimilarity index with the Adonis function from the Vegan package (Table 1A) in R. The principal coordinate analysis, using Bray-Curtis dissimilarity index, was carried out with the ordinate function of the phyloseq package in R. Alpha diversity, beta diversity and relative abundances of the taxa in each sample were plotted with ggplot2 package (Table 1A) in R.

A phylogenetic tree was obtained from the 16S rRNA sequencing data. First, all ASVs in influent, effluent and dried sludge samples were aligned using MUSCLE (Table 1B). The alignment file was then used for generating a tree with FastTree (Table 1B).

VALIDATION OF PRIMERS FOR EPICPCR

In order to select the target genes for epicPCR, the efficacy of 14 primer sets for 10 different ARGs of interest was tested (Appendix 1: ermB, ermF, tnpA-04, intI1_3, intI1_4, blaIMP, blaCTX-M-04, blaNDM, tetG, sul1, sul2_1, sul2_2, sul2_3, sul2_4). The primers were selected based on previous literature and they were adapted for being used in epicPCR by adding the required overhangs. To validate their suitability for epicPCR, the same PCRs that are performed in epicPCR were carried out with each primer set. Differently from epicPCR, for testing purposes DNA extracted from influent and effluent wastewaters was used as template rather than isolated cells. The samples from which the

DNA was extracted were previously collected on 04/03/2019 from Viikinmäki WWTP and their DNA extracted with DNeasy PowerWater Kit (QIAGEN).

Fusion PCR was performed in quadruplicate in 25 µL reactions. Each reaction contained 2 µL of template DNA and the reagents reported in Table 2A. The PCR program for fusion PCR was the one originally described (Spencer et al., 2016). The four replicates for each reaction were pooled and the product was purified with Monarch PCR & DNA Cleanup kit (New England Biolabs). The same kit was used to purify pooled PCR products after each PCR when testing the primers.

Since the method was developed, the epicPCR protocol was optimized by adding an extra blocking PCR step before nested PCR (unpublished). The only primers used in blocking PCR are the blocking primers, preventing the amplification of unfused products. This extra step seems to improve the selection of the target product in the following nested PCR, so it was included in this study. To evaluate whether the use of blocking primers was effective for each primer set, after fusion PCR a reaction with blocking primers and one without blocking primers were run.

In the first case, a blocking PCR was carried out in 25 µL reactions in two replicates, using 2 µL of purified fusion PCR product as the template and the reagents indicated in Table 2B. The PCR program for blocking PCR was the following: beginning temperature 98°C for 30 seconds; 30 cycles of denaturation at 98°C for 10 seconds, annealing at 55°C for 30 seconds, extension at 72°C for 30 seconds; final extension at 72°C for 5 minutes. Two microliters of the blocking PCR product were then used as template for nested PCR, which had a reaction volume of 25 µL, two replicates per sample, and included a negative control. The reagents used are shown in Table 2C. The PCR program for nested PCR was the one originally presented (Spencer et al., 2016), with 55°C as the annealing temperature.

In the second case, no blocking PCR was performed. Nested PCR was analogous to the one described for the first case, but it did not include the blocking primers and the template DNA was the purified fusion PCR product.

Nested PCR replicates were pooled and the products were visualized with E-Gel™ EX Agarose Gels, 2% (Invitrogen). Products obtained from the same PCR mix with influent and effluent as template were purified, pooled and sequenced with Illumina MiSeq sequencing platform at the Institute of Biotechnology of University of Helsinki.

Table 2. Reagents used in epicPCR

A) Fusion PCR	
Reagent	**Final concentration**
GC buffer (New England Biolabs)	1X
$MgCl_2$ (New England Biolabs)	1 mM
dNTPs (BioNordika)	0.25 mM
Phusion Hot Start Flex DNA polymerase (New England Biolabs)	0.16 U/µL
Forward primer (F1) (Appendix 1)	1 µM
Reverse primer (pH') (Appendix 1)	1 µM
Linker primer (R1-F2') (Appendix 1)	0.01 µM

B) Blocking PCR	
Reagent	**Final concentration**
GC buffer (Thermo Fisher Scientific)	1X
dNTPs (BioNordika)	0.2 mM
Phusion polymerase (Thermo Fisher Scientific)	0.02 U/µL
U519_blockF (Appendix 1)	3.2 µM
U519_blockR (Appendix 1)	3.2 µM

C) Nested PCR	
Reagent	**Final concentration**
GC buffer (Thermo Fisher Scientific)	1X
dNTPs (BioNordika)	0.2 mM
Phusion polymerase (Thermo Fisher Scientific)	0.02 U/µL
U519_blockF (Appendix 1)	0.32 µM
U519_blockR (Appendix 1)	0.32 µM
Forward nested primer (F3_TS or F1_TS) (Appendix 1)	0.3 µM
Reverse 16S primer (Illum_785R1-4) (Appendix 1)	0.3 µM

ANALYSIS OF PRIMER TESTING PCR PRODUCTS

After sequencing of the PCR products obtained from the testing of epicPCR primers, Illumina adapters were removed from the reads by using cutadapt (Table 1B). The paired-end reads were then joined with PEAR (Table 1B) using default options. Cutadapt was used to remove the 16S end primer and to filter out the reads with quality lower than PHRED=20 (99% accuracy). At each step, the quality of the reads was checked with FastQC (Table 1B). The fastq files containing the merged and filtered reads were converted into fasta format. The reads obtained from different samples but targeting the same gene were combined into one file. For each target gene, the forward primer sequence was used for extracting the reads that contained the target gene part by running cutadapt. The reads were then split into the target gene part and the 16S rRNA gene part using cutadapt and the bridge primer sequence for each target gene. To ensure that the target gene primers in fact amplified the gene of interest, the part of the read that supposedly corresponded to the target gene part was aligned to a nucleotide database with BLAST (Table 1B). For primers targeting ARGs, the part of each read corresponding to the target gene was aligned to the resfinder database (Zankari et al., 2012)

by using blastn with default parameters. For reads obtained with primers blaIMP and blaCTX-M-04 only, the blastn parameters for short reads were used by specifying blastn-short in the -task option. For primers targeting MGEs (intI1_3, intI1_4, tnpA-04), the part of each read corresponding to the target gene was aligned to a MGEs database (Pärnänen et al., 2018).

GENE DETECTION

To further direct the choice of target genes for epicPCR, the following genes were targeted with PCR for detection in the same samples that would then be used in epicPCR: *tetM, strB, ermB, ermF, sul1, sul2, int1, bla$_{IMP}$, bla$_{CTX-M-04}$, bla$_{NDM}$, tetG*. The templates used were mixes of all DNA extraction replicates from each sample type, diluted 1:10. The primers used were F1 and R2-F1´ for each gene (Appendix 1). Only sul2_3 primers were used for detecting *sul2*, while both intI-1_3 and intI-1_4 were used for detecting *int1*. The reagents that were used are the following: 1X Phusion GC buffer (Thermo Fisher Scientific), 0.2 mM dNTPs (BioNordika), 0.02 U/µL Phusion DNA polymerase (Thermo Fisher Scientific). The final concentration of both primers was 0.5 µM. The amount of template DNA added to each reaction was 1 µL. The reaction volume was 25 µL and for each primer set four reactions were performed, with different templates: one with influent DNA, one with effluent DNA, one with dried sludge DNA and one negative control. The PCR program was the following: initial denaturation at 98°C for 30 seconds; 30 cycles of denaturation at 98°C for 10 seconds, annealing at the optimal temperature for each primer set for 30 seconds, extension at 72°C for 30 seconds; final extension at 72°C for 5 minutes. The annealing temperature was 59°C for tetM, strB and ermB primers, 64°C for int1 (with both primer sets intI1_3 and intI1_4), blaIMP, blaCTX-M-04 and blaNDM primers, 63°C for tetG, ermF and sul1 primers, 68°C for sul2 primer (sul2_3). The products were visualized with E-Gel™ EX Agarose Gels, 2% (Invitrogen).

EPICPCR

Considering the results obtained from the validation of epicPCR primers and the detection of different ARGs in the samples, the following genes were chosen to be targeted in epicPCR: *ermB, ermF, bla$_{IMP}$, bla$_{NDM}$, tetM, strB, sul1, sul2*.

The cells that were extracted from the samples and stored in glycerol were used as template for epicPCR. The cell suspensions were thawed, centrifuged for 1 minute at 12 000 x *g* and resuspended in 30 µL of PCR-quality water. The bead formation protocol was adapted from Spencer et al., 2016, by introducing the same modifications presented in Hultman et al., 2018. After polymerization, the frequency of beads containing no cells, one cell, and two or more cells was evaluated by staining the cells with SYBR™ Green II DNA Gel Stain (Invitrogen) and inspecting a fraction of the beads with a compound microscope (Zeiss Axioskop 2 plus).

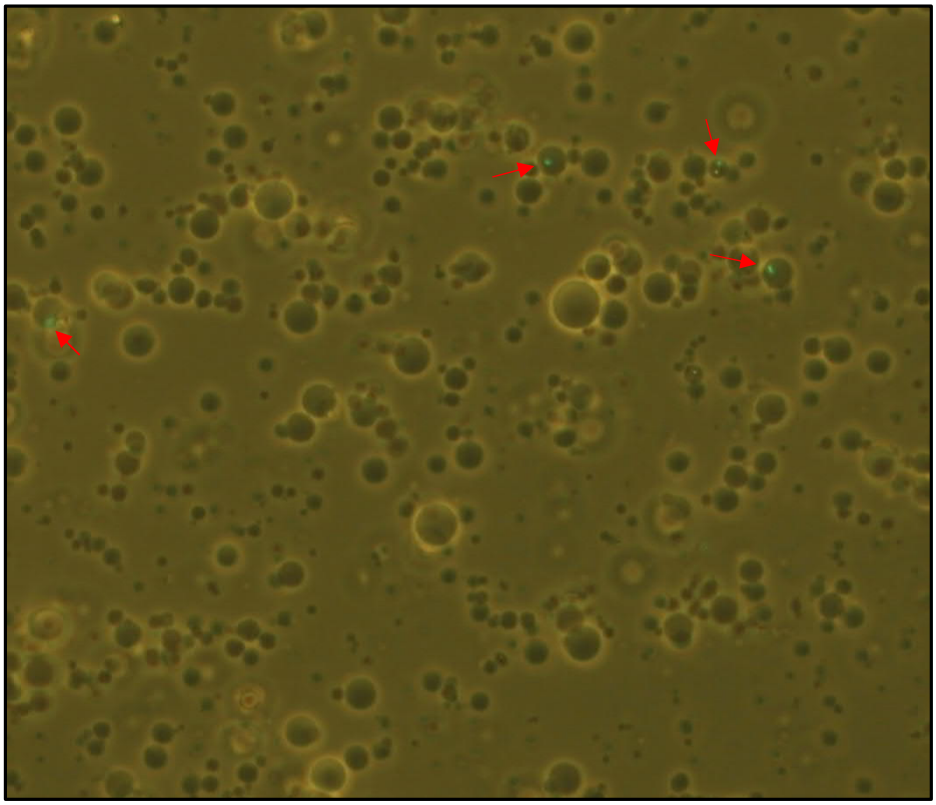


Figure 3 Polyacrylamide beads to be used in epicPCR. The figure shows beads obained from influent water collected on the third sampling day (i3). Most of the beads are empty, and a few beads contain one isolated bacterial cell, stained with green fluorescence and indicated by the red arrows. No beads contain more than one cell, which is ideal.

For each sample, several cell dilutions were used as starting material for bead formation, and the one with the best prevalence of beads containing one cell was chosen for fusion PCR (Figure 3).

For fusion PCR, the protocol presented in Spencer et al., 2016, was followed, with the same modifications presented in Hultman et al., 2018. The reagents are shown in Table 2A. No lysis step was carried out before fusion PCR, and no EDTA was added to the pooled sample aliquots after the PCR as in Spencer et al., 2016. In the breaking of ABIL emulsions phase, diethyl ether was added as the first thing, with no prior centrifugation step. After the first diethyl ether extraction, 50 µL of water was added to the sample before repeating the extraction. Differently from how it was described in Spencer et al., 2016, Monarch PCR & DNA Cleanup Kit (New England Biolabs) was used for purifying the fusion products. The same kit was used for PCR product purification in the following steps, after pooling PCR replicates.

As explained in the "Validation of primers for epicPCR" section, an additional blocking PCR was performed before nested PCR. The reaction setup was the same described previously, with reagents shown in Table 2B, but with a reaction volume of 50 µL.

Nested PCR was performed as in Spencer et al., 2016, using 2 µL of undiluted purified blocking PCR product as a template for each 25 µL reaction. The annealing temperature was 55°C. The reagents are

presented in Table 2C. A negative control was included in each reaction. The primers used in nested PCR already contained Illumina TrueSeq adapter overhangs, so the additional Illumina PCR described in Spencer et al., 2016, was not performed. The replicates for each sample were pooled and the products visualized with E-Gel™ EX Agarose Gels, 2% (Invitrogen).

If no product was visible in the agarose gel, or the negative control had a visible band, epicPCR was repeated for that sample. For the following samples epicPCR was repeated a second time: bla_{IMP} influent, *strB* influent, all effluent samples, *ermB* sludge, bla_{IMP} sludge, *ermF* sludge 3, *sul1* sludge 2 and 3, bla_{NDM} sludge 1 and 2. The following samples were also repeated a third time: *ermB* effluent 3, *sul1* sludge 2 and 3, bla_{NDM} sludge 1 and 2. These repeated reactions were considered as technical replicates.

Nested PCR products were purified and sequenced with Illumina MiSeq sequencing platform at the Institute of Biotechnology of University of Helsinki. As *tetM* and *sul2* did not give any visible PCR product, no products for these genes were sequenced. For the other genes (*ermB, ermF,* bla_{IMP}*,* bla_{NDM}*, sul1, strB*) all technical replicates and nested PCR negative controls were sequenced.

EPICPCR DATA ANALYSIS

The reads obtained from the sequencing of epicPCR purified products were filtered, merged and split into target gene and 16S rRNA gene parts as previously described in the "Analysis of primer testing PCR products" section of this work. The target gene parts were also aligned to the resfinder database (Zankari et al., 2012), using BLAST (Table 1B), in the same way as for the primer testing PCR products.

The 16S rRNA gene parts extracted from all reads were renamed to retain the information on the ARG they were linked to, and then combined into one file. Chimeras were removed from the combined file with the VSEARCH tool (Table 1B). First, chimeras were detected de novo, and afterwards a reference-based chimera detection was carried out. The Silva "Gold" reference database, a curated version of the Silva database (Quast et al., 2013) designed for chimera removal (available at https://www.mothur.org/w/images/f/f1/) was used as reference. Mothur (Table 1B) was used to assign taxonomy to all the non-chimeric 16S rRNA sequences that were linked to a target ARG. The mothur classify.seqs command was used with default parameters, and the reference database was Silva version 132 (Quast et al., 2013). Next, sequences for which no taxonomy was assigned were removed in mothur with the remove.lineage command by specifying taxon=unknown.

At this point, samples that had less 16S rRNA reads than their corresponding negative controls were discarded. Also, when a biological replicate had less than 1% of the reads of the other replicates, it

was not considered. For those samples for which epicPCR was repeated more than once, the best technical replicate in terms of number of reads and negative control purity was chosen (Appendix 2). The information contained in the output taxonomy table was processed with R (Table 1A). The host range of each ARG was analysed independently from the others, by extracting from the taxonomy table the 16S rRNA gene reads that were linked to that ARG. The host range of the target ARGs was evaluated at the genus level. Because of the qualitative nature of epicPCR, the number of reads assigned to each genus in a sample was not considered, and the results were reduced to presence/absence matrices. In R, presence of a gene in a genus in a sample type was evaluated by removing duplicate genera in a sample, merging the lists of genera from the three biological replicates of the same sample type and counting the occurrence of each genus in the list.

An ARG was considered to be present in a genus in a sample type (e.g. influent) when the link was found in all three biological replicates for that sample type (3/3, e.g. influent 1, influent 2 and influent 3, sampled on three different dates). This aimed at reducing the number of false positives resulting from such cases when more than one cell is randomly encapsulated in the same polyacrylamide bead in epicPCR. When for a sample type one of the biological replicates did not have a good number of 16S rRNA reads, a gene was considered present in a genus when found in both the other biological replicates (2/2). When two or all three of the biological replicates had less than 20 reads in all technical replicates, presence/absence was not evaluated for that gene in that sample type.

For bla_{IMP}, due to the fact that not all target gene parts of the reads were aligning to the correct gene, 16S rRNA gene sequences that were linked to something other than bla_{IMP} had to be removed in R. After removing such sequences, the analysis was carried out in the same way as for the other genes.

Reads that were not classified at the genus level, but were assigned only to upper taxonomic levels, were removed from the presence/absence tables. The presence/absence tables showing the host range of each ARG in the three sample types were plotted using ggplot2 package (Table 1A) in R, after combining them with the abundance values for each genus obtained from the 16S rRNA gene sequencing data.

The presence/absence data for each gene and the genus abundance values were also visualized together with the previously made phylogenetic tree of the whole WWTP bacterial community using Anvi'o (Table 1B) interactive interface. The figure was then refined using Inkscape, an open-source vector graphics editor (available from http://inkscape.org/).

Negative controls were kept separate in the analysis, and their 16S rRNA gene content was evaluated separately. Tables containing unique genera were created in R for each negative control and then inspected manually.

RESULTS

MICROBIAL COMMUNITY COMPOSITION

The composition and diversity of the microbial community of Viikinmäki WWTP were investigated by amplifying and sequencing the 16S rRNA gene from influent water, effluent water and dried sludge samples. A taxonomic classification was assigned to each detected amplicon sequence variant (ASV), and the relative abundance of different taxa was compared between samples. The diversity of the microbial community was estimated both within a sample (alpha diversity) and between samples (beta diversity).

The alpha diversity, representing the diversity of the microbial community in each sample, was calculated with Shannon diversity index. The bacterial community of effluent water was less diverse compared to the ones of influent and sludge, and it also exhibited a higher variance between the three samples collected on different days. The diversity measure was more similar for influent water and dried sludge, with dried sludge being the most diverse (Figure 4).

The beta diversity of the samples was calculated with a principal coordinate analysis using Bray-Curtis dissimilarity index. The microbial community of effluent water was found to be significantly different from the ones of influent water and dried sludge, based on the permutational multivariate analysis of their variance ($p\text{-value}_{\text{influent-effluent}} = 0.01667$, $R^2_{\text{influent-effluent}} = 0.86333$, $p\text{-value}_{\text{effluent-sludge}} = 0.01667$, $R^2_{\text{effluent-sludge}} = 0.9169$) (Figure 5). The microbial communities of influent water and dried sludge appear to be distinct based on the principal coordinate analysis plot (Figure 5), but their difference was not found to be significant ($p\text{-value}_{\text{influent-sludge}} = 0.06667$, $R^2_{\text{influent-sludge}} = 0.95228$). However, because of the small sample size ($n = 3$), the statistical power of these comparisons was limited.

Based on the abundance of amplicon sequence variants (ASVs), the most abundant bacterial order in influent wastewater was Campylobacteriales, to which 25%-35% of the reads from the three replicates were assigned (Figure 6). Other abundant orders were Clostridiales (13%-14%), Betaproteobacteriales (9%-15%), Lactobacillales (9%-14%) and Pseudomonadales (10%-11%). At the genus level, the most abundant taxon in influent samples was *Arcobacter* (26%-36%), followed by *Acinetobacter* (7%-9%), *Trichococcus* (6%-9%) and *Flavobacterium* (4%-7%) (Figure 7).

In effluent water, Betaproteobacteriales was the most abundant order, with 55%-80% of the reads being assigned to it (Figure 6). The genus *Methylotenera*, belonging to the order Betaproteobacteriales, was the most abundant genus in effluent samples, with a relative abundance of 44%-73% (Figure 7).

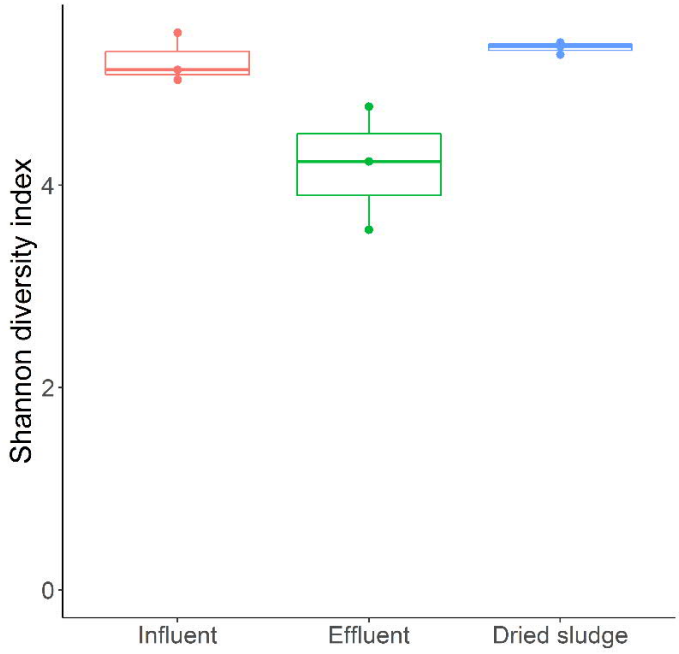


Figure 4 Alpha diversity of the microbial community of influent water, effluent water and dried sludge, calculated with Shannon index. The diversity of each sample type in the three sampling days is represented with a dot. The effluent community is the least diverse, while dried sludge samples had the highest diversity in their microbial community.

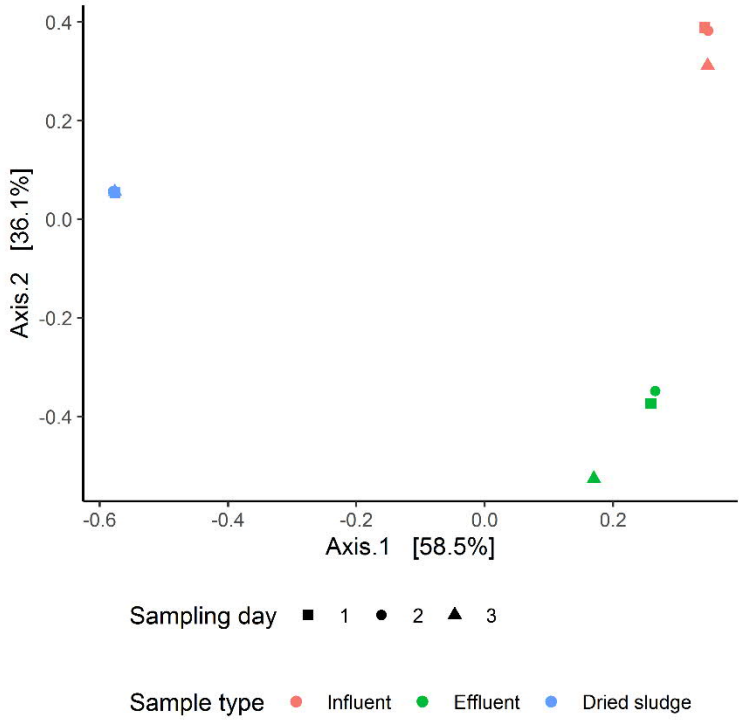


Figure 5 Principal coordinate analysis plot calculated with Bray-Curtis dissimilarity index from ASVs in influent, effluent and dried sludge samples. The two dimensions that were plotted explain almost all of the variation between sample types (94,6%). Different sample types cluster separately, but influent and dried sludge were not found to be significantly different. On the other hand, the microbial community of effluent water was significantly different from the ones of influent water and dried sludge.

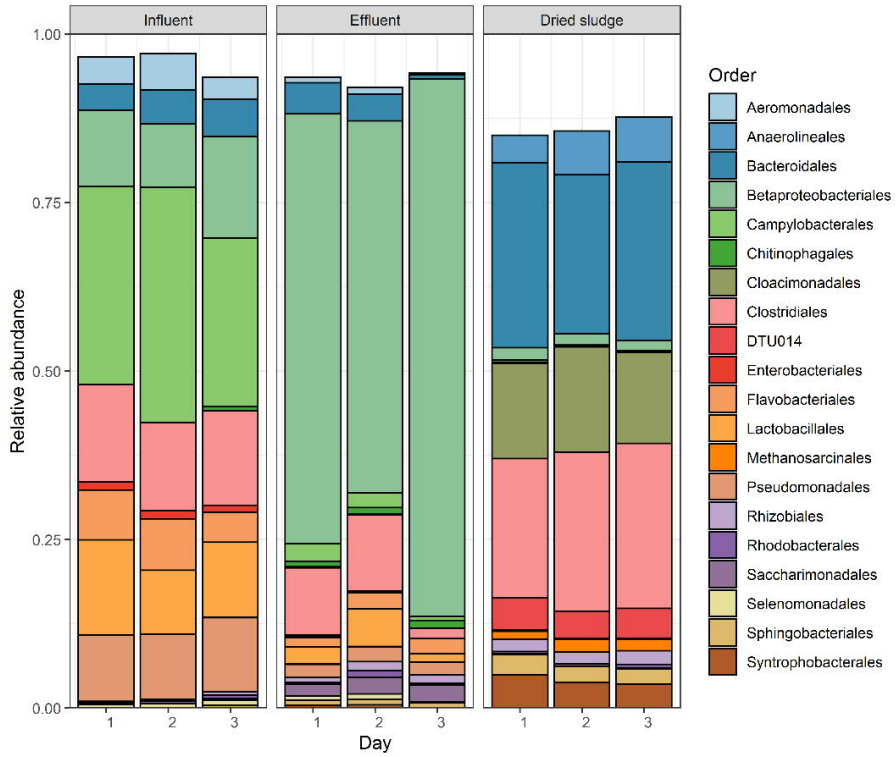


Figure 6 Twenty most abundant orders in influent, effluent and dried sludge samples collected on three different days, based on taxonomic classification of the 16S rRNA gene sequences obtained from them.

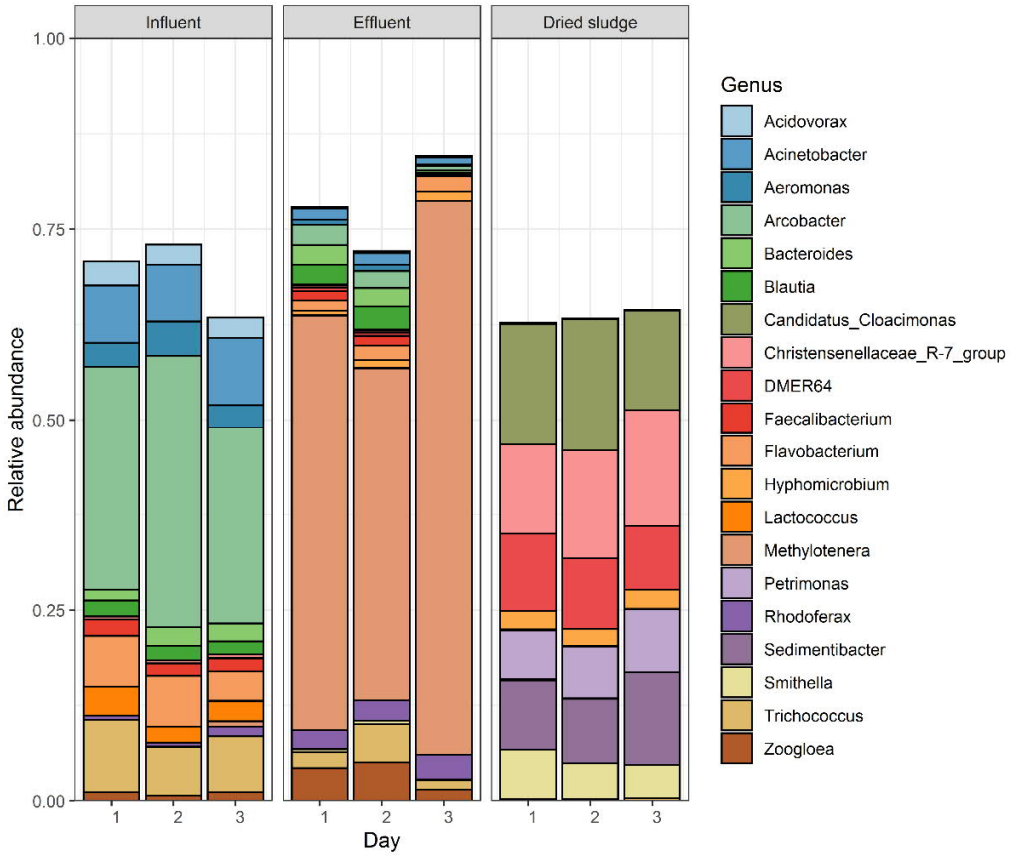


Figure 7 Twenty most abundant genera in influent, effluent and dried sludge samples collected on three different days, based on taxonomic classification of the 16S rRNA gene sequences obtained from them.

Clostridiales was the second most abundant order in effluent water, with 10%-11% of the reads from sampling days 1 and 2 being assigned to it (Figure 6). In the third effluent sample only 1.4% of the reads were assigned to Clostridiales. The relative abundance of other orders was below 5%. The second most abundant genus in the effluent was *Zoogloea*, whose relative abundance was 1%-5% (Figure 7).

In dried sludge, Bacteroidales, Clostridiales and Cloacimonadales were the most abundant orders, with a proportion of reads assigned to them equal to 24%-27%, 21%-25% and 13%-16% respectively, followed by Anaerolineales (4%-7%), DTU014 (4%-5%) and Syntrophobacterales (4%-5%) (Figure 6). Among the bacterial genera, Candidatus *Cloacimonas* was the most abundant in sludge samples, with a relative abundance of 13%-17%. Other abundant genera were *Christensenellaceae R-7 group* (12%-15%), *Sedimentibacter* (8%-12%), and *DMER64* (8%-10%) (Figure 7). The reads assigned to the genus *Petrimonas* were 6%-8% of the total, and 4%-6% were assigned to *Smithella*.

Overall, the main bacterial groups in the influent, effluent and dried sludge communities differ substantially.

Both the DNA extraction from the samples and the 16S rRNA gene PCR included negative controls, which were sequenced and analysed analogously to the samples. All negative controls contained no reads after all filtering steps and merging, except for three of them. The absolute number of reads assigned to each of the ASVs in samples and negative controls was compared. After comparing the abundance of each ASV that was present in negative controls with that in the respective samples, no sequences were removed from the samples (Figure 8). Contaminants were present in very low amounts in negative controls and they were frequent in real samples, suggesting that their presence in negative controls would be the result of cross-contamination coming from the samples.

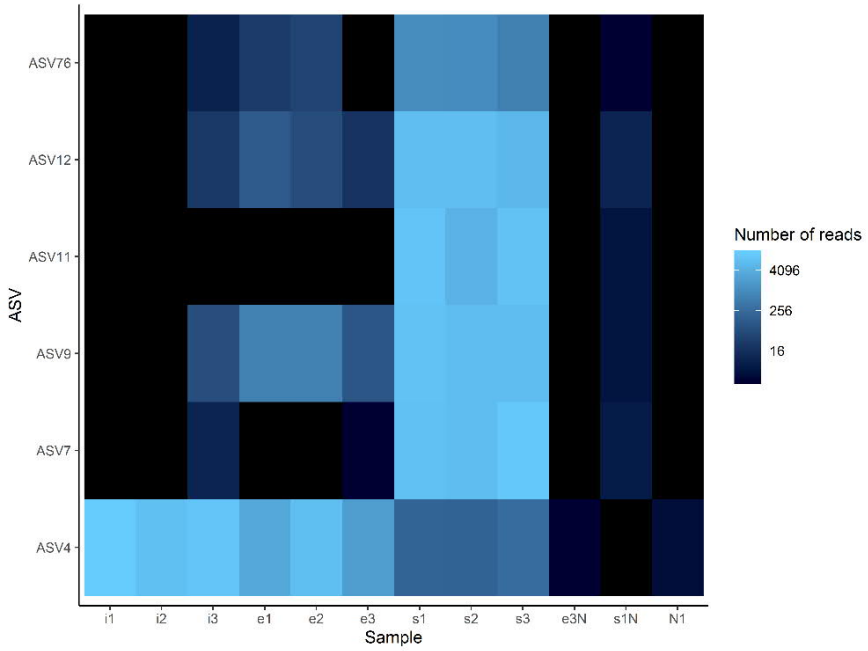


Figure 8 Heatmap of the number of reads present in the negative controls and all samples of 16S rRNA gene sequencing, after all filtering steps. e3N is the DNA extraction negative control of e3, and it contains 1000-fold less copies of the same ASV (ASV4). s1N is the DNA extraction negative control of s1 (sludge, day 1) and it contains ASVs 7, 9, 11, 12 and 76 in 1000-fold less copies than its respective sample s1. N1 is the PCR negative control for i1, i2, i3, s1, s2, s3 (influent day 1, 2, 3; sludge day 1, 2, 3) and it contains ASV4 in 100-fold less copies than sludge samples s1, s2 and s3 and in 1000-fold less copies than influent samples i1, i2 and i3.

CHOICE OF GENES FOR EPICPCR: EFFICACY OF THE TESTED PRIMERS AND PRESENCE OF ARGS IN THE WWTP

In order to choose the genes to be targeted with epicPCR, primers for each shortlisted gene were adapted for epicPCR use and tested with three PCR steps. The presence of the genes was also detected in the samples with PCR.

The efficacy of the tested primer sets in amplifying the target gene fused with 16S rRNA gene is shown in Table 3. Primers targeting *ermB, ermF, bla*$_{NDM}$*, sul1, sul2* (sul2_3 and sul2_4) and *int1* (intI1_3 and intI1_4) gave as a result high numbers of reads that were all aligning to the correct target gene. For blaIMP primers, 11884 reads gave hits aligning to *bla*$_{IMP}$ gene (99,93% of the total number of reads), while 8 gave hits aligning to other resistance genes (0,07% of the total number of reads). TetG and tnpA-04 primers gave products that aligned to the correct target gene, but the number of reads obtained for these genes was small, suggesting that the primers might not be optimally efficient. Only 1 read from sul2_1 primers was aligned to a resfinder entry, and it corresponded to the correct gene. Sul2_2 primers did not result in any product aligning to resfinder entries. For blaCTX-M-04 primers, only 6 reads were aligned to resfinder entries, 2 of which aligned to *bla*$_{CTX\text{-}M\text{-}59}$ gene and the

remaining 4 to different resistance genes. Overall, ermB, ermF, blaNDM, sul1, sul2_3, sul2_4, intI1_3, intI1_4 and blaIMP primers were considered suitable for being used in epicPCR.

ErmB, ermF, tetM, strB, sul1, sul2, tetG and *int1* were detected in all three sample types (Figure 9 A, B, D, E). Bla_{IMP}, bla_{NDM} and $bla_{CTX\text{-}M\text{-}04}$ were not detected in any of the samples (Figure 9 C, D).

Based on these results, *tetG, tnpA-04* and $bla_{CTX\text{-}M}$ genes were excluded from the list of genes to be targeted with epicPCR, because their primers did not give high numbers of reads aligning to the correct gene. *Int1* was also excluded from the epicPCR analysis, in order to restrict the study to ARGs only. Bla_{IMP} and bla_{NDM} were chosen to be targeted with epicPCR even though they were not detected in the samples, because their primers were efficient and because of their clinical relevance. For targeting *sul2*, the sul2_3 primer set was chosen. The other genes of choice were *sul1, ermB* and *ermF*, for which primers were efficiently validated, and *tetM* and *strB*, whose primers had already been tested previously.

Table 3. Efficacy of the tested epicPCR primers in targeting the correct gene.

PRIMER SET	BLAST HIT	NUMBER OF READS	PERCENTAGE OF READS ALIGNING TO THE CORRECT GENE
ermB	erm(B)	9629	100%
ermF	erm(F)	13389	100%
blaIMP	blaIMP vanXmurFvanKWI tet(V) tet(51) mph(A) fosB1 cmlV blaVIM-47 blaOXA-299	11892 1 1 1 1 1 1 1 1	99,93%
ndm-1	blaNDM	11401	100%
blaCTX-M-04	blaCTX-M-59_1 erm(V) dfrA17 aph(3')-Id aadA5	2 1 1 1 1	33,33%
sul1	sul1	18318	100%
sul2_1	sul2	1	100%
sul2_2			
sul2_3	sul2	7064	100%
sul2_4	sul2	9965	100%
tetG	tet(G)	228	100%
tnpA-04	2228_tnpA	196	100%
intI1_3	269_intI1	10405	100%
intI1_4	419_intI1	14577	100%

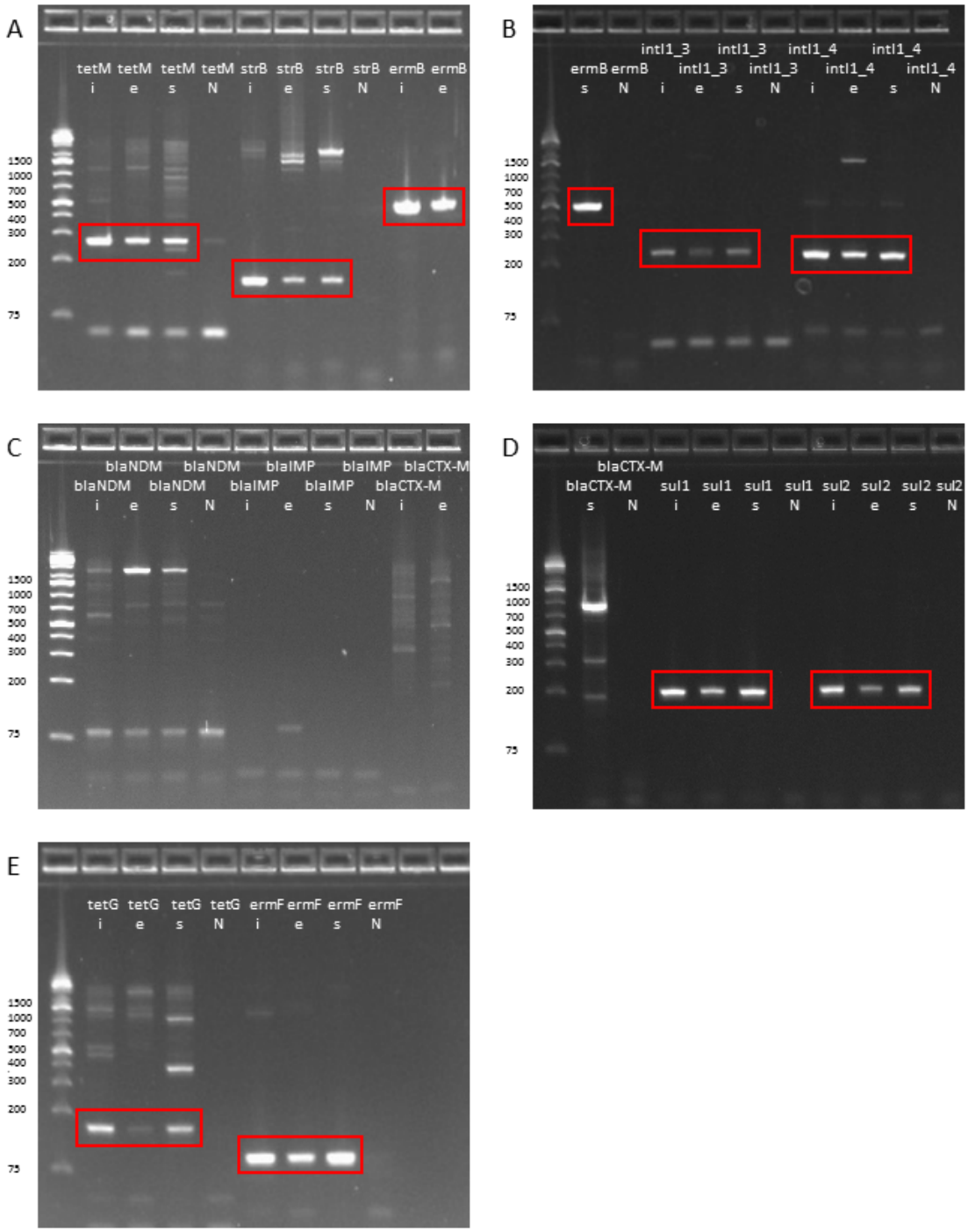


Figure 9 Detection of ARGs and MGEs in influent, effluent and dried sludge. For each tested gene, one PCR negative control was included (i = influent water, e = effluent water, s = dried sludge, N = negative control). The following ARGs were detected in all sample types: A) *tetM* (expected product size: 256 bp), *strB* (expected product size: 135 bp); A-B) *ermB* (expected product size: 499 bp); D) *sul1* (expected product size: 185 bp), *sul2* (expected product size: 190 bp); E) *tetG* (expected product size: 140 bp), *ermF* (expected product size: 79 bp). C) The ARGs bla_{NDM} (expected product size: 176 bp), bla_{IMP} (expected product size: 72 bp) and $bla_{CTX\text{-}M\text{-}04}$ (C-D, expected product size: 62 bp) were not detected in any of the samples. B) The integrase 1 gene (*int1*) was detected in all samples with two different primer sets (intI1_3 and intI1_4, expected product size for both: 217 bp).

HOST RANGE OF TARGET ARGS

The host range of a series of ARGs was investigated by using Emulsion, Paired Isolation and Concatenation PCR (epicPCR) on influent water, effluent water and dried sludge samples. The eight genes that were chosen to be targeted with epicPCR were *ermB, ermF*, bla_{IMP}, bla_{NDM}, *tetM, strB, sul1, sul2*. Of these, only six gave PCR products for at least one sample in a quantity sufficient to be visualized with gel electrophoresis: *ermB, ermF,* bla_{IMP}, bla_{NDM}, *strB, sul1*. Therefore, products for *tetM* and *sul2* were not sequenced and their host range was not evaluated. After sequencing, not all genes gave a sufficient number of reads in all samples for their host range to be resolved. For this reason, the host range of *strB* was analysed only in influent water samples, and that of bla_{IMP} only in influent and effluent water (Appendix 2). EpicPCR targeting bla_{NDM} resulted in an inadequate number of reads for one influent and one sludge sample, and that was also the case for *sul1* in one sludge sample (Appendix 2). Among all samples, the average number of 16S rRNA reads passing all filtering steps and being assigned a taxonomic classification was 34 884.

The β-lactamase gene bla_{IMP} was found in 19 different genera in the influent and 11 in the effluent, while its host range was not resolved in dried sludge samples (Figure 10). The gene was linked to some abundant genera, namely *Acinetobacter* and *Arcobacter* in influent water and *Methylotenera* in effluent water, but also several low abundance genera. In addition, bla_{IMP} was found in two genera in effluent, *Methylophilus* and *Parabacteroides*, that were not hosting it in influent. Neither of these two genera showed a considerable increase in frequency when comparing effluent to influent water.

The other β-lactamase gene under investigation, bla_{NDM}, was not detected in dried sludge. In influent water, its host genera were *Acinetobacter, Aeromonas, Arcobacter, Comamonas, Klebsiella, Methylotenera, MM1, Pseudomonas* and *Stenotrophomonas*, among which *Acinetobacter* and *Arcobacter* were especially abundant in the sample (Figure 10). The only genus to which the gene was associated both in influent and effluent water was *Methylotenera*, while it was hosted by the genera *Dechloromonas* and *Methylophilus* only in effluent water. Neither *Dechloromonas* nor *Methylophilus* were among the most abundant genera in effluent.

ErmB was associated to 10 bacterial genera in influent, of which 4 also carried it in effluent and 4 hosted it both in influent and sludge (Figure 10). *Acinetobacter, Methylophilus* and *Methylotenera* were carrying *ermB* in all three sample types, while *Arcobacter* only hosted it in influent and sludge and *Escherichia-Shigella* only in influent and effluent. In the genera *Bacteroides* and *Zoogloea*, the gene was only detected in effluent water, as the association was not found in influent and sludge. Similarly, *ermB* was found to be hosted by *Tolumonas* only in dried sludge, despite the low abundance of the genus in that sample type.

ErmF was, among the genes that were investigated, the one with the broadest host range. It was found in 73 bacterial genera in influent, 23 in effluent and 34 in dried sludge (Figure 10). The host range included abundant genera but also several rare ones. The genera in which the gene was found in all three sample types were *Acinetobacter, Aeromonas, Aquaspirillum, Arcobacter, Flavobacterium, Methylophilus, Methylotenera, MM1, Pseudomonas, Stenotrophomonas, Tolumonas* and *Zoogloea*. In several cases, *ermF* was associated to a genus in dried sludge but not in influent. Such genera were *Bergeriella, C39, H1, Kinneretia, Parabacteroides, Petrimonas, Proteiniphilum, Simplicispira* and *SN8*. Except for *Petrimonas*, none of these genera was among the most abundant ones in sludge. *Parabacteroides, Simplicispira* and *SN8* were carrying *ermF* both in sludge and effluent, albeit not in influent, while the gene was associated with *Smithella* in effluent only.

Due to the low number of reads that were obtained with *strB* primers both from effluent and from dried sludge samples, the host range of this gene was only investigated for influent water samples. In influent water, *strB* was found to be associated with 19 genera, including *Acinetobacter, Aeromonas, Arcobacter, Comamonas, Dechloromonas, Pseudomonas* and *Zoogloea* (Figure 10). *Acinetobacter* and *Arcobacter* were both abundant in influent, while the other genera had a lower relative frequency in the sewage water community.

The gene conferring resistance to sulphonamide, *sul1*, was harboured by 19 genera in influent water, 20 genera in effluent water and 10 genera in dried sludge (Figure 10). Similarly to the other genes, it was associated with abundant genera such as *Arcobacter* and *Methylotenera*, but also with rare genera. In *Acinetobacter, Aeromonas, Arcobacter, Delftia, Methylophilus, Methylotenera, Pseudomonas, Stenotrophomonas* and *Zooglea*, *sul1* was found in each of the three sample types. In the following genera, *sul1* was detected only in effluent water: *Ideonella, MM1, Sulfurospirillum, Tibeticola, Tolumonas* and *Variovorax.* The gene was also found in effluent and sludge, but not influent, associated with *Aquaspirillum.*

Acinetobacter, Aeromonas, Arcobacter, Comamonas, Pseudomonas and *Stenotrophomonas* were carrying all the genes in influent water. *Acinetobacter* also hosted bla_{IMP}, *ermB, ermF* and *sul1* in effluent, and *ermB, ermF* and *sul1* in sludge. *Arcobacter* was one of the host genera for bla_{IMP}, *ermF* and *sul1* in effluent, and for bla_{IMP}, *ermB, ermF* and *sul1* in sludge. *Methylophilus* and *Methylotenera* harboured all the target ARGs in effluent water, excluding *strB*, for which not enough sequences were obtained to evaluate its host range in effluent. No genus was found to carry all the target genes in dried sludge.

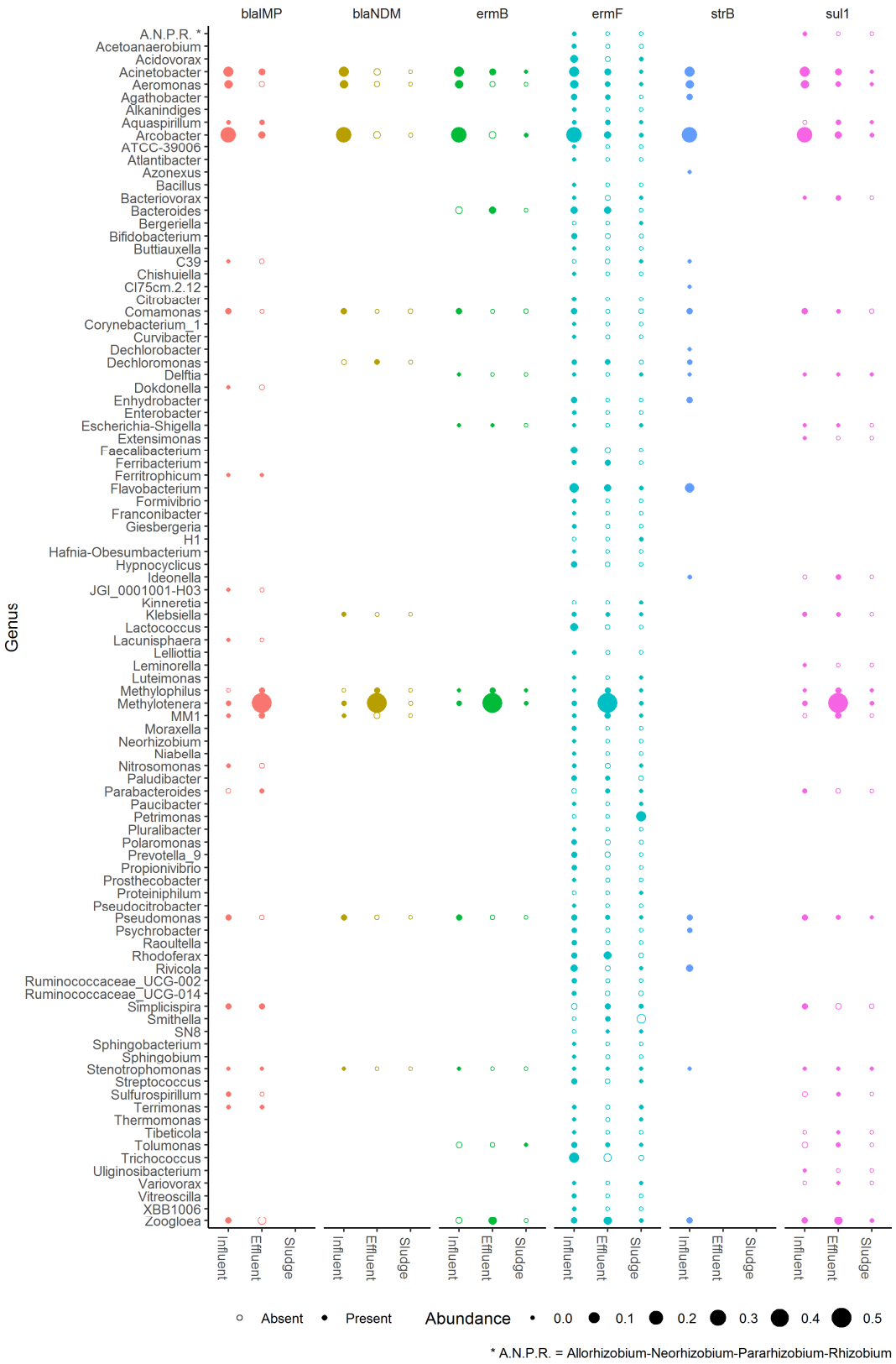


Figure 10 Host range of *bla*$_{IMP}$, *bla*$_{NDM}$, *ermB*, *ermF*, *strB* and *sul1* genes in influent water, effluent water and dried sludge. Genera that were found to be linked to at least one of the ARGs are displayed. Coloured dots indicate that the gene was detected in that genus, while empty dots indicate that the gene was absent. Dots are not displayed when the genus was not found to be linked to the gene in any sample type. The host range of *strB* in effluent and sludge and of *bla*$_{IMP}$ in sludge was not analysed due to the low number of sequences. The size of each dot represents the relative abundance of the genus in the sample, based on 16S rRNA gene sequencing.

The presence of ARGs in bacterial genera and the relative abundance of the genera in the three samples was also visualized in relation to the phylogenetic tree of the bacterial community, in order to evaluate whether certain taxonomic groups were more prone to carrying the target ARGs (Figure 11). The great majority of the ARG-carrying genera belonged to the phylum Proteobacteria. Also, this phylum included almost all the genera carrying all the target ARGs. The only genus that carried all the target ARGs and was not assigned to Proteobacteria was *Arcobacter*, which belongs to Epsilonbacteraeota. Other phyla included genera carrying only up to three of the target ARGs. Among the target ARGs, *ermF* had the broadest host range not just in terms of number of hosts, but also in the diversity of bacteria carrying it (Figure 11).

When combining the epicPCR results with the abundance data, not all genera that were found to be linked to some ARG were also detected with the sequencing of the 16S rRNA gene from the total community. For instance, this was the case for *Azonexus, Bacillus, CI75cm.2.12, Delftia* and *Pseudocitrobacter*.

Negative controls from each reaction, that were filtered and analysed in the same way as the rest of the samples but did not have technical replicates for presence/absence calling, generally contained low numbers of reads and hosts. In cases when a sample contained less 16S rRNA reads than its respective negative control, it was not considered in the analysis (Appendix 2). As negative controls are only introduced in nested PCR, they do not give information on the contaminants that might come from the work environment and reagents, but rather they show traces of cross-contamination between PCR reactions. Some, but not all of the genera that were ubiquitous in the samples, such as *Acinetobacter, Arcobacter, Methylotenera* and *Stenotrophomonas*, were also found in some negative controls. However, because they were present in low abundance in the negative controls compared to their respective samples and they were most likely sign of cross-contamination between biological replicates, these genera were not removed from the analysis.

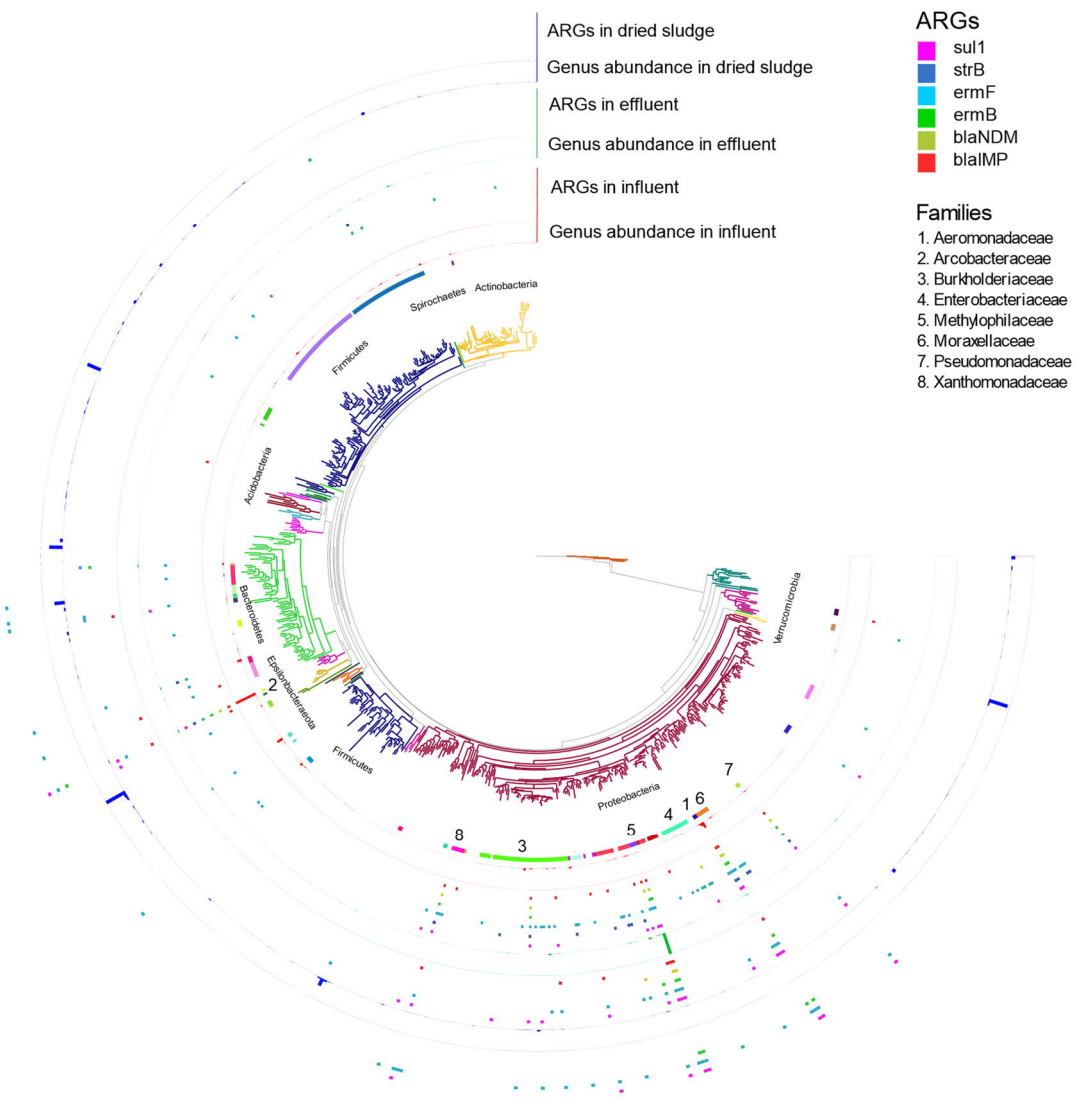


Figure 11 Abundance and presence of ARGs in bacterial genera of a WWTP. The tree was constructed with the 16S rRNA gene sequences obtained from the whole microbial community of influent water, effluent water and dried sludge combined. Each branch represents a genus and it is coloured based on the phylum it belongs to. The name of the phylum was reported only for those that were found to host at least one ARG based on epicPCR results. In the first layer, families harbouring at least one ARG are represented with colours. The most relevant families in terms of detected multi-resistance are numbered and reported in the legend. For each sample type (influent water, effluent water, dried sludge) the abundance of each genus is represented on a logarithmic scale, followed by the ARG presence in each genus.

DISCUSSION

VARIATIONS IN THE MICROBIAL COMMUNITY COMPOSITION

The abundance of different bacterial groups in influent water, effluent water and dried sludge samples was investigated by sequencing the 16S rRNA gene from the whole microbial community. Considering the variations in frequency of different taxa is essential for better interpreting possible changes in the host range of ARGs as detected with epicPCR.

The microbial community of effluent water was significantly different from that of influent water and dried sludge. Also, it was the one with the lowest alpha diversity. The reduced richness of the effluent water community is likely to result from the selection pressures that exist in the WWTP environment, such as the ones resulting from aeration, temperature variations, and the addition of compounds like methanol. The difference between dried sludge and influent water communities, on the other hand, was not statistically significant. The two communities appear nonetheless distinct, and it is not to exclude that reason behind the lack of statistical significance of their differences might lie in the limited sample size, with only three replicates per type.

During the treatment, some microbial groups can be enriched due to the process itself. This seems to be the case for methylotrophs, such as the family Methylophilaceae, whose relative abundance in effluent water was 43%-73%, exhibiting a sharp increase compared to its frequency in influent water, which was lower than 0.7% in all replicates. In dried sludge, Methylophilaceae also represented less than 1% of the microbial families. The increase in their frequency is most likely due to the addition of methanol, which Methylophilaceae use as source of carbon and energy, during the denitrification step of the treatment, which is carried out after the removal of the sludge and thus only affects the effluent water composition.

The residual sludge, before being dried and released, goes through an anaerobic digestion step. This has an impact on the relative abundance of the microbial community members. The most abundant genera in sludge, Candidatus *Cloacimonas, Christensenellaceae R-7 group, Sedimentibacter,* and *DMER64,* were rare in influent, with frequencies below 1%. In effluent water, the frequencies of these genera were also below 1%. Since they are present in very low abundance in influent and they are not enriched in the effluent, these genera which are abundant in sludge are probably enriched during the sludge digestion and its processing before release.

Overall, the results obtained from the analysis of the microbial community composition of influent and effluent water were in line with the ones presented in the most recent study reporting the microbial community composition of Viikinmäki influent and effluent waters (Hultman et al., 2018). Small differences in the results from the two studies can be attributed to the different sampling time, which

was March in this study and September in the previous one (Hultman et al., 2018). Sampling in spring or in autumn implies differences in terms of water temperature and nutrient content, which can both affect the microbial community composition. Another difference between the two studies was the pipeline that was used for analysing the 16S rRNA gene sequences. In Hultman et al., 2018, the sequences were clustered in OTUs with 97% identity, which were then classified. In this study, instead of clustering the sequences into OTUs, a taxonomic classification was assigned to each amplicon sequence variant (ASV) with the DADA2 package, which allows for a higher resolution (Callahan et al., 2016). In addition, in this study, release 132 of the Silva database (Quast et al., 2013) was used, instead of release 128 as in Hultman et al., 2018.

Relying on different databases, as well as using different tools for data analysis, poses an obstacle to the comparability of the results. One example of this can be seen in the effluent community composition reported in the two studies. In this work, the most abundant bacterial order in effluent water collected from Viikinmäki WWTP was Betaproteobacteriales, while in Hultman et al., 2018, Methylophilales was found to be the most frequent, followed by Neisseriales. However, this difference is only apparent, since there was a drastic change in the classification of Proteobacteria in the Silva 132 release (Quast et al., 2013). In particular, Betaproteobacteria, once a class, became an order, Betaproteobacteriales. As a consequence, both the orders Methylophilales and Neisseriales became families (Methylophilaceae and Neisseriaceae) within the order Betaproteobacteriales. This is to represent how even minor changes in a pipeline, such as the utilization of a different database, can make comparisons less straight-forward.

Investigating how the relative abundance of different bacterial groups changes during the treatment process is an important prerequisite to the study of variations in the host range of ARGs. Indeed, when detecting an ARG associated to a host in effluent water or dried sludge but not in influent water, it might be tempting to conclude that the gene has been actively selected, or that it has been acquired by that host in a HGT event occurred inside the treatment plant. However, failing to detect a gene in a certain host in the influent, while detecting it in effluent or sludge, can simply be due to a low abundance of that host in the influent, causing it to be undetected with epicPCR. In addition, during the treatment process, the sewage water community also interacts with the plant resident community, in particular with that of activated sludge. Therefore, effluent water and dried sludge could contain antibiotic resistant bacteria that were not at all present in the influent, because they acquired them from the WWTP community. To discriminate the different sources of ARGs in effluent water and dried sludge, it would be very informative to study antibiotic resistant bacteria in activated sludge and WWTP biofilms.

HOST RANGE OF ARGS AS DETECTED WITH EPICPCR

The host range of six ARGs (*bla$_{IMP}$, bla$_{NDM}$, ermB, ermF, strB, sul1*) in samples collected from Viikinmäki WWTP was resolved by applying epicPCR. This method was used to target ARGs on wastewater samples only once before, with *tetM, bla$_{OXA-58}$, qacEΔ1* and *int1* being the genes of choice (Hultman et al., 2018). EpicPCR enables a high throughput linking of a target gene with the bacteria harbouring it in a microbial community, bypassing any culturing steps. For this reason, it is very promising for studying the distribution of ARGs in complex environmental communities, like the ones associated with wastewaters.

Overall, this study identified differences in the host range of the target genes in influent, effluent and dried sludge which might provide evidence for the occurrence of HGT inside the WWTP. Unfortunately, not all the targeted genes gave sufficient results for investigating their host range. This was the case for the genes *tetM* and *sul2*, even though they were both detected in all the samples.

Among the target genes that were chosen, the New Delhi metallo-β-lactamase gene *bla$_{NDM}$* was of great interest in this study. Occasionally detected in wastewaters (Hendriksen et al., 2019), it arose only in recent times, and rapidly spread across the globe (Johnson and Woodford, 2013). This gene is of high clinical relevance because it confers resistance to carbapenems, which are crucial last resort antibiotics that are used to treat multi-resistant infections. In this study, *bla$_{NDM}$* was not detected in dried sludge, while it was associated with 12 genera in total in influent and effluent waters. In influent water it was found to be associated with previously reported *bla$_{NDM}$*-carrying genera, specifically *Acinetobacter, Aeromonas, Comamonas, Escherichia, Klebsiella, Pseudomonas, Stenotrophomonas* (Johnson and Woodford, 2013; Ou et al., 2014). None of these genera was carrying *bla$_{NDM}$* in effluent waters. *Arcobacter, Dechloromonas, Methylophilus, Methylotenera* and *MM1* were detected for the first time as hosts of *bla$_{NDM}$*. Interestingly, *Dechloromonas* and *Methylophilus* were associated with the gene only in effluent water.

The second β-lactamase coding gene under investigation, *bla$_{IMP}$*, is also relevant in the medical field, as it confers resistance to carbapenems and other β-lactam antibiotics. The host range of *bla$_{IMP}$* could not be resolved in dried sludge samples because of the high level of cross-contamination of the epicPCR products (Appendix 2). In most genera, *bla$_{IMP}$* was found in influent and occasionally also in effluent waters. Only in two genera, *Parabacteroides* and *Methylophilus*, it was found in effluent but not influent. Neither of these genera was remarkably more abundant in effluent water than in influent water.

Two MLSB resistance genes were also included in the study, *ermB* and *ermF*. The first, *ermB*, is one of the most abundant ARGs in wastewaters worldwide (Hendriksen et al., 2019), but its host range was not found to be especially broad in Viikinmäki WWTP. *Bacteroides* and *Zoogloea* were only

carrying the gene in effluent water, while *Tolumonas* only hosted it in dried sludge. *Zoogloea* was more abundant in effluent water than in influent water, while in the case of *Bacteroides* and *Tolumonas* the abundance of the genus was unvaried or decreased after the treatment.

The second MLSB resistance gene, *ermF*, had a remarkably broad host range compared to the other genes that were examined, especially in influent and dried sludge samples. The possibility of detecting more hosts could be related to the fact that more reads were obtained from the sequencing of PCR products for *ermF* compared to the other genes, which allowed for a higher resolution of its host range (Appendix 2). Its previously reported enrichment in dried sludge (Karkman et al., 2016) does not seem to be fully explained by an enrichment of the bacteria carrying it. Among the genera carrying *ermF* both in influent water and dried sludge, the only one that had a remarkable increase in frequency during the process was *Petrimonas*. Notably, in nine cases the gene was reported to be present in a genus in sludge but not in influent.

The sulphonamide resistance gene *sul1* is one of the most abundant ARGs in sewage water, and it is characterized by a broad host range also due to its frequent association with the class I integron (Gillings et al., 2008). In this study, *sul1* was detected in 26 bacterial genera. In seven genera (*Aquaspirillum, Ideonella, MM1, Sulfurospirillum, Tibeticola*, *Tolumonas*, *Variovorax*) the gene was found in effluent or both in effluent and dried sludge, but not in influent. None of these genera exhibited a sharp increase in frequency during the treatment process.

The aminoglycoside resistance gene *strB*, which is one of the predominant genes in sewage waters (Hendriksen et al., 2019), is known to have a broad host range because it is often localized on MGEs (Sundin and Bender, 1996). In this study, its host range could not be resolved in effluent and dried sludge samples because of the low number of reads obtained from them. For this reason, possible changes in the range of bacteria carrying it across samples cannot be discussed. However, in influent samples *strB* was found in 19 genera, including rare genera and genera that were not associated with other target ARGs (*Azonexus, CI75cm.2.12*).

All the reported cases in which a gene was found in a host in effluent water or dried sludge, but not in influent water, imply that HGT events might have occurred. However, it is also necessary to consider that, when a gene is found to be linked to a genus in effluent or sludge, but not in influent, a potential explanation can be given by an increase in the relative abundance of the genus, which would make the association between host and gene easier to detect. This would suggest that the claim according to which the detection of an ARG in a host only in effluent or sludge might be the result of HGT is stronger when the host is not substantially more abundant in these samples than in influent. However, even in such cases, a possible reason for its association with an ARG only in effluent and sludge could be that selection occurred at the sub-genus or even sub-species level. For instance, the

genus *Methylophilus* was here found to harbour bla_{IMP} and bla_{NDM} only in effluent water, but its relative abundance was similar in influent and effluent. This observation could be explained with HGT events, but it could also result from the fact that a hypothetical *Methylophilus* species or subspecies was already carrying the genes in the influent while being too rare to be detected, and it became more abundant in the effluent, which caused it to be detected with epicPCR. In this scenario, the other *Methylophilus* species or subspecies would become less abundant during the treatment, so the abundance of the genus would not visibly change. Looking at sequence variants rather than taxonomic annotations when comparing epicPCR results with 16S rRNA sequencing results would partially prevent this issue when interpreting results. In any case, all possible scenarios need to be considered before claiming that ARGs have been transferred to different hosts in the WWTP, as the results might be a consequence of detection limits of epicPCR. Unfortunately, methods for real-time detection of HGT events inside the WWTP are not yet established, which restricts the possibilities of confirming our results.

Looking at the host range of all target genes, we can identify some bacterial genera that carry several resistance determinants. *Acinetobacter* was carrying all the genes in influent water, of which four were also associated with it in effluent water and three in dried sludge. This genus was already known to carry the targeted ARGs (Johnson and Woodford, 2013; Lee et al., 2017; Roberts et al., 1999; Roberts, 2008; Zhao and Hu, 2011). Several *Acinetobacter* species are pathogenic. In particular, *Acinetobacter baumannii* is the most important cause of nosocomial infections globally, and several multi-resistant strains of it have been isolated (Lee et al., 2017). The presence in wastewaters of *Acinetobacter* species carrying all the ARGs that were analysed in this work is therefore consistent with previous literature. Even though no signs of acquisition of such genes by *Acinetobacter* during the process were seen, the detection of antibiotic resistant *Acinetobacter* species in effluent and sludge confirms that WWTPs could release multi-resistant pathogens into the environment.

Arcobacter was also harbouring all the target ARGs in influent water. *Arcobacter* species, which are among the most abundant in WWTPs (Mclellan et al., 2015; Wu et al., 2019), are found in numerous and diverse environments (Ferreira et al., 2019). Some of them are known human or animal pathogens (Ferreira et al., 2019), which makes the detection of resistance genes in *Arcobacter* especially interesting. In meta-analytic studies, *Arcobacter* species have been found to have resistant phenotypes towards macrolide, aminoglycoside and β-lactam antibiotics (Ferreira et al., 2019). However, the genes that were targeted in this work were not directly identified in *Arcobacter* in previous studies. Interestingly, the *Arcobacter* genus was also found to harbour all the ARGs that were previously targeted with epicPCR on wastewater samples from Viikinmäki WWTP (Hultman et al., 2018). Only

a small fraction of *Arcobacter* isolates carries plasmids (Douidah et al., 2014), which is why the detection of such a wide range of ARGs in this genus is perplexing.

Other genera carrying all ARGs in influent samples were *Aeromonas, Comamonas, Pseudomonas* and *Stenotrophomonas*. Each of these include human pathogens, such as *Aeromonas hydrophila*, *Comamonas testosteroni, Pseudomonas aeruginosa* and *Stenotrophomonas maltophilia*. Based on the results presented here, the wastewater treatment process appears to be effective in removing resistant bacteria belonging to these genera, with the exception of resistance genes *ermF* and *sul1*.

A puzzling result which was obtained in this study was the association of all target genes, with the exception of *strB*, with methylotrophic genera, namely *Methylotenera* and *Methylophilus*. The association was particularly evident for effluent water samples, in which both genera were found to host bla_{IMP}, bla_{NDM}, *ermB, ermF* and *sul1*. The Methylophilaceae family is greatly enriched in effluent water, and its increase in frequency is due to the wastewater treatment process. These bacteria are in fact used in biofilters for denitrification of sewage water, so their growth is actively promoted by adding methanol to the water to be treated (Rissanen et al., 2017). No members of the Methylophilaceae family are known to be human or animal pathogens (Shen et al., 2019), which is possibly why their antibiotic resistance profile has not yet been studied. Methylophilaceae have been shown to be enriched in the presence of antibiotics on greenhouse lettuce, but the reasons underlying this observed behaviour are unknown (Shen et al., 2019). The MGEs content of this bacterial family is also unexplored. However, in accordance with the results presented in this work, members of the Methylophilaceae family were previously found to harbour *int1* and *tetM* genes in effluent water from Viikinmäki WWTP, but not in influent water (Hultman et al., 2018).

The detection of the target ARGs in bacteria that are highly abundant in the microbial community and that are not known to carry them, like in the case of *Arcobacter* in influent water samples and Methylophilaceae in effluent waters, is enigmatic. Theoretically, it might be easier to detect a gene in a genus if that genus is abundant in the sample. However, the detection of ARGs in low abundance genera shows that epicPCR is sensitive and that genes harboured in rare bacteria can still be effectively amplified. One point that also needs to be considered when interpreting these results is that more abundant bacteria are more likely to cause false positives in epicPCR. As a matter of fact, the chances of a genus not carrying an ARG being encapsulated in the same bead with an ARG-carrying cell, resulting in an incorrect pairing of the ARG with the 16S rRNA gene from a non-host, are increased when the genus is abundant in the community. This does not justify the detection of ARGs in these genera also in samples where they were not so abundant, for example in *Methylotenera* in influent water. However, the possibility of cross-contamination between different reaction also needs to be considered. The limits of epicPCR as a method are further discussed in the next section.

To ensure that the detected genera really carry multiple resistance genes, further research is required. Metagenomic analyses could be carried out for the same samples, but linking an ARG and its hosts with this method is challenging. Another option would be the isolation of the species of interest from wastewaters, followed by testing of their susceptibility to different antibiotics or sequencing of their whole genome to identify resistance determinants. Alternatively, microfluidic devices could be used to sort single cells from the whole microbial community into plates. Afterwards, each individual cell could be parallelly targeted for the amplification of both the 16S rRNA gene and an ARG of interest. By sequencing the 16S rRNA gene of the isolates in which the target ARG was amplified, it would be possible to identify the bacterial groups hosting it.

If the results presented here were to be confirmed, they would imply that *Arcobacter* species, of which some are pathogenic, harbour ARGs more frequently than previously thought. The detection of genes conferring resistance to last resort antibiotics, such as bla_{IMP} and bla_{NDM}, in *Arcobacter*, would be especially alarming from the clinical point of view. The validation of these results would also suggest that members of the Methylophilaceae family are hosts of several ARGs. Even though this family does not include any pathogenic species, its abundance in effluent waters potentially makes it a key actor in the dissemination of resistance determinants in the receiving environments, where they might be transferred to other bacteria.

This study found dried sludge to include less multi-resistant genera compared to influent water and effluent water. Also, only for three of the six genes (*ermB, ermF* and *sul1*) at least one host was identified in dried sludge samples. This might suggest that the role of this sample type in disseminating resistance genes to the environment is more limited compared to effluent water. However, failing to detect certain genes in dried sludge with epicPCR may also be a consequence of the higher complexity in applying the method to this sample type. Indeed, cells are not so easily isolated from sludge samples as from water samples, due to the presence of aggregates and solid particles. In addition, this might cause a greater retainment of PCR inhibitors, which would interfere with epicPCR. It is also worth noting that the samples were collected in March, and the gene abundance in sludge was seen to be at its lowest in the spring season (Karkman et al., 2016). Repeating the study with samples collected throughout the year might give more comprehensive results.

One further possible source of error in this study was the fact that, while the reverse primer that was targeting the 16S rRNA gene was the same for 16S rRNA gene sequencing and epicPCR (785R, Appendix 1), two different forward primers were used for the two studies. 341F primer (Appendix 1) was used for 16S rRNA gene sequencing, whereas 536F (the reverse complement of 519R, Appendix 1) primer was the part of the epicPCR bridge primer that was acting as a forward primer for 16S

rRNA gene in epicPCR, as previously described (Hultman et al., 2018). The primer pair 341F-785R amplifies the variable regions V3 and V4 of the 16S rRNA gene, allowing a better taxonomic resolution compared to the primer pair 536F-785R, that only amplifies the V4 region. Because the primers that were used were targeting different regions of the same gene, it is not possible to directly compare the sequences obtained from them. In other words, 16S rRNA gene sequencing data cannot be matched with the 16S rRNA part obtained from sequencing of epicPCR product. Instead, these two different datasets can only be compared in terms of taxonomic annotations. This approach is not ideal, because it fails to detect sequence variants and it is not successful in such cases when the taxonomic annotation of a sequence has not been carried out up to the lowest level. Also, universal 16S rRNA gene primers might be slightly biased, failing to amplify all 16S rRNA genes from a community. This could explain why some genera that were found with epicPCR (e.g. *Azonexus*, which was linked to *strB)* were not detected in the 16S rRNA gene sequencing data, that supposedly represents the whole community. To avoid these complications in merging data from 16S rRNA gene sequencing and epicPCR, the same primers should be used for the two analyses. This would allow to evaluate the abundance of a taxon carrying a target gene in epicPCR by comparing directly its sequence to the one obtained by 16S rRNA gene sequencing, rather than comparing taxonomic assignations, therefore making the results more accurate.

LIMITS OF EPICPCR

EpicPCR has been developed very recently and it has not yet been used extensively. Due to this, the method in general is still subject to optimization and currently has some potential sources of error. In the light of the work that was presented here, there are several points of the method that could be targeted to make it more effective, more versatile, and more accurate.

First, in this work and the previous published ones (Cairns et al., 2018; Hultman et al., 2018; Qin et al., 2019; Spencer et al., 2016), the portion of 16S rRNA gene that was amplified with epicPCR only covered the V4 variable region of the gene. This stretch of DNA is too short for the taxonomic annotation to reach the species level. By sequencing a longer part of the 16S rRNA gene, ideally the whole gene, it would be possible to make the taxonomy annotation more accurate and therefore more informative. For instance, looking at epicPCR results at the species level could show whether the gene is carried by a pathogen or not, which would be especially useful when a genus includes both pathogenic and non-pathogenic species, of which *Arcobacter* is an example. Also, the closer we could look at the sequence variants of the whole 16S rRNA genes, the more accurate the evaluation of presence/absence of a gene and the differences between samples would be. In this work, the sequence

of the entire 16S rRNA gene would have helped to verify whether a species that was carrying a gene in different sample types was the same, implying the flow of the antibiotic resistant bacteria through the WWTP, or whether there were different variants of it, which would suggest possible HGT events occurring in the plant. So far, the possibility of linking longer 16S rRNA gene parts to the target gene in epicPCR has been limited by the sequencing technology that was used. Indeed, sequencing platforms that produce short reads cannot be used for long amplicons. In the future, epicPCR products could be sequenced with other technologies that produce long reads. In addition to allowing for a higher resolution on the 16S rRNA gene, being able to sequence longer amplicons would also open the possibility for amplifying longer stretches of the target gene, which would be useful for tracking its transmission and carrying out phylogenetic studies.

Another point that could be improved in the epicPCR protocol concerns negative controls. Fused products are artificial constructs that do not exist in nature. Therefore, any product that is present in the negative controls of the final nested PCR and that contains both the target gene part and the 16S part is likely to result from cross-contamination between samples, rather than from contaminants in the reagents or in the working environment. Since nested PCR uses PCR amplicons as template, preparing several reactions simultaneously makes them highly subject to cross-contamination. More careful and sterile procedures for handling PCR products could reduce the levels of cross-contamination. Moreover, no negative controls are included in the first steps of the epicPCR protocol. Introducing a negative control in the bead formation phase, which would consist in making polyacrylamide beads containing water instead of bacterial cells, would be useful for detecting possible bacterial contaminants that might be carried over during the whole downstream process, and cannot be detected by only having a negative control in nested PCR.

In addition, the formation of polyacrylamide beads in an emulsion results in beads of different sizes. This can cause uneven amplification of the DNA of different cells during fusion PCR, as cells that are captured in larger beads also have a larger quantity of PCR reagents available, which can result in more PCR product. Besides preventing the method from being quantitative, this issue can make it harder to detect bacteria that carry the target gene but are randomly incapsulated into smaller beads. Using techniques that lead to the generation of beads of constant size, for example with microfluidic-based devices, would bring an improvement in the accuracy of the method.

Lastly, the false positive rate associated with epicPCR is still unknown. False positives would arise from cross-contamination of PCR products from different samples, but also in those cases when more than one cell is randomly encapsulated in the same bead. This would lead to the incorrect linking of the target gene with bacteria that were not in reality carrying it. So far, technical or biological replicates were used as a proof of presence of a gene in a host: if the association was found in several

replicates, it was believed not to be caused by random encapsulation of that bacterial group with others carrying the gene. However, as previously mentioned for cases such as *Arcobacter* and *Methylotenera* in this work, it is plausible that more abundant taxa would have higher chances of causing false positives. Although the presence of a gene in a host can be then verified with other tools, the false positive rate associated with epicPCR should be addressed in order to make the method more valid.

CONCLUSIONS

In this study, the host range of six ARGs (*bla$_{IMP}$*, *bla$_{NDM}$*, *ermB*, *ermF*, *strB*, *sul1*) in influent, effluent and dried sludge samples collected from Viikinmäki WWTP was successfully resolved by applying epicPCR. Variations in the microbial community composition were also studied by sequencing the total 16S rRNA gene content of the three sample types, and used to complement the information obtained with epicPCR.

The results presented here suggest that WWTPs are indeed hotspots for the transmission of ARGs between different hosts. By applying epicPCR, several putative HGT events that might have occurred inside the WWTP were identified.

Another finding of this study was that the target ARGs, most notably *bla$_{NDM}$*, are found also in bacterial genera that were not previously known to harbour them. This might partly depend on the fact that the detection of ARGs in isolates is usually carried out for pathogens, with the result that ARGs are more rarely identified in environmental bacteria.

Overall, all targeted ARGs were detected both in abundant and in rare bacterial genera. Notably, in water samples, both influent and effluent, all bacterial groups that were highly abundant were also found to carry multiple ARGs. This also applied to genera that had not been reported to show antibiotic resistant phenotypes before, specifically the members of the Methylophilaceae family.

Further research is needed to confirm that the bacteria that were here indicated as ARG-carriers truly harbour the target genes. Even though epicPCR is an extremely promising tool for unravelling the host range of functional genes in microbial communities, the method itself still includes some sources of error and can be further optimized. At this point, it is crucial to identify weaknesses and limits of the method in order to give a more accurate interpretation of the results obtained by applying it.

ACKNOWLEDGEMENTS

I wish to acknowledge the Helsinki Region Environmental Services Authority (HSY) for collecting and providing the samples that were used in this work, and the CSC—IT Center for Science, Finland, for providing the computational resources that were needed for the data analysis.

I would like to thank my supervisor Antti Karkman for guiding me through all stages of this project, and for all the help and feedback he gave me whenever needed. I am also grateful to my PI Marko Virta for giving me the possibility to carry out this project and his valuable advices. Finally, I wish to express my gratitude to my fellow lab mates. Thanks for all your support, it has been great to work together on our respective thesis projects.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W., and Lipman, D.J. (1990). Basic Local Alignment Search Tool. J. Mol. Biol. *215,* 403-410.

Andrews, S. (2011). FastQC: A quality control tool for high throughput sequence data. Available online at: bioinformatics.babraham.ac.uk/projects/fastqc/ [last accessed on 20/11/2019].

Bengtsson-Palme, J., Hammarén, R., Pal, C., Östman, M., Björlenius, B., Flach, C., Fick, J., Kristiansson, E., Tysklind, M., and Larsson, D.G.J. (2016). Elucidating selection processes for antibiotic resistance in sewage treatment plants using metagenomics. Sci. Total Environ. *572,* 697-712.

Bonomo, R.A. (2017). β-Lactamases: A Focus on Current Challenges. Cold Spring Harb. Perspect. Med. *7*, a025239.

Cacace, D., Fatta-Kassinos, D., Manaia, C.M., Cytryn, E., Kreuzinger, N., Rizzo, L., Karaolia, P., Schwartz, T., Alexander, J., Merlin, C., *et al.* (2019). Antibiotic resistance genes in treated wastewater and in the receiving water bodies: A pan-European survey of urban settings. Water Res. *162,* 320-330.

Cairns, J., Ruokolainen, L., Hultman, J., Tamminen, M., Virta, M., and Hiltunen, T. (2018). Ecology determines how low antibiotic concentration impacts community composition and horizontal transfer of resistance genes. Communications Biology *1*, 1-8.

Callahan, B.J., Mcmurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods *13,* 581-583.

Chopra, I., and Roberts, M. (2001). Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance. Microbiology and Molecular Biology Reviews *65,* 232–260.

Collignon, P., Beggs, J.J., Walsh, T.R., Gandra, S., and Laxminarayan, R. (2018). Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. The Lancet Planetary Health *2,* e398-e405.

Crofts, T.S., Gasparrini, A.J., and Dantas, G. (2017). Next-generation approaches to understand and combat the antibiotic resistome. Nature Reviews Microbiology *15,* 422-434.

Davies, J. (2007). Microbes have the last word. A drastic re-evaluation of antimicrobial treatment is needed to overcome the threat of antibiotic-resistant bacteria. EMBO Rep. *8,* 616-621.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. Journal of Vegetation Science *14,* 927-930.

Douidah, L., De Zutter, L., Van Nieuwerburgh, F., Deforce, D., Ingmer, H., Vandenberg, O., Van Den Abeele, A.M., and Houf, K. (2014). Presence and analysis of plasmids in human and animal associated Arcobacter species. PLOS One *9*, e85487

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32,* 1792-1797.

Edwards, U., Rogall, T., Blöcker, H., Emde, M., and Böttger, E.C. (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. Nucleic Acids Res. *17*, 7843-7853.

Eren, A.M., Esen, ÖC., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ *3*, e1319.

Ferreira, S., Luís, Â, Oleastro, M., Pereira, L., and Domingues, F.C. (2019). A meta-analytic perspective on Arcobacter spp. antibiotic resistance. Journal of Global Antimicrobial Resistance *16,* 130-139.

Gillings, M., Boucher, Y., Labbate, M., Holmes, A., Krishnan, S., Holley, M., and Stokes, H.W. (2008). The Evolution of Class 1 Integrons and the Rise of Antibiotic Resistance. The Journal of Bacteriology *190,* 5095-5100.

Guo, B., Liu, C., Gibson, C., and Frigon, D. (2019). Wastewater microbial community structure and functional traits change over short timescales. Sci. Total Environ. *662,* 779-785.

Hendriksen, R.S., Munk, P., Njage, P., Bunnik, B.V., Mcnally, L., Lukjancenko, O., Röder, T., Nieuwenhuijse, D., Pedersen, S.K., Kjeldgaard, J., *et al.* (2019). Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nature Communications *10,* 1-12.

Herlemann, D.P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. The ISME Journal *5,* 1571–1579.

Hultman, J., Tamminen, M., Parnanen, K., Cairns, J., Karkman, A., and Virta, M. (2018). Host range of antibiotic resistance genes in wastewater treatment plant influent and effluent. FEMS Microbiol. Ecol. *94,* fiy038.

Johnson, A.P., and Woodford, N. (2013). Global spread of antibiotic resistance: the example of New Delhi metallo-beta-lactamase (NDM)-mediated carbapenem resistance. J. Med. Microbiol. *62,* 499-513.

Karkman, A., Do, T.T., Walsh, F., and Virta, M.. (2018). Antibiotic-Resistance Genes in Waste Water. Trends Microbiol. *26,* 220-228.

Karkman, A., Johnson, T.A., Lyra, C., Stedtfeld, R.D., Tamminen, M., Tiedje, J.M., Virta, M., and Topp, E. (2016). High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. FEMS Microbiol. Ecol. *92,* fiw014.

Karkman, A., Pärnänen, K., and Larsson, D.G.J. (2019). Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. Nature Communications *10,* 80.

Laht, M., Karkman, A., Voolaid, V., Ritz, C., Tenson, T., Virta, M., and Kisand, V. (2014). Abundances of tetracycline, sulphonamide and beta-lactam antibiotic resistance genes in conventional wastewater treatment plants (WWTPs) with different waste load. PLOS One *9,* e103705.

Lee, C., Lee, J.H., Park, M., Park, K.S., Bae, I.K., Kim, Y.B., Cha, C., Jeong, B.C., and Lee, S.H. (2017). Biology of Acinetobacter baumannii: Pathogenesis, Antibiotic Resistance Mechanisms, and Prospective Treatment Options. Frontiers in Cellular and Infection Microbiology *7,* 55.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17,* 10-12.

Mclellan, S., Fisher, J., and Newton, R. (2015). The microbiome of urban waters. International Microbiology *18,* 141-149.

McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS One *8,* e61217.

O'Neill, J. (2016). Tackling drug-resistant infections globally: final report and recommendations. AMR Review, available at https://amr-review.org/sites/default/files/160518_Final%20paper_with%20cover.pdf [last accessed: 20/11/2019]

O'Neill, J. (2014). Antimicrobial resistance: Tackling a crisis for the health and wealth of nations. AMR Review, available at https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf [last accessed: 20/11/2019]

Ou, W., Cui, L., Li, Y., Zheng, B., and Lv, Y. (2014). Epidemiological characteristics of blaNDM-1 in Enterobacteriaceae and the Acinetobacter calcoaceticus-Acinetobacter baumannii complex in China from 2011 to 2012. PLOS One *9,* e113852.

Pal, C., Bengtsson-Palme, J., Kristiansson, E., and Larsson, D.G.J. (2016). The structure and diversity of human, animal and environmental resistomes. Microbiome *4,* 54.

Pärnänen, K., Narciso-Da-Rocha, C., Kneis, D., Berendonk, T., Cacace, D., Do, T., Elpers, C., Fatta-Kassinos, D., Henriques, I., Jaeger, T., *et al.* (2019). Antibiotic resistance in European wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence. Science Advances *5,* eaau9124.

Pärnänen, K., Karkman, A., Hultman, J., Lyra, C., Bengtsson-Palme, J., Larsson, D.G.J., Rautava, S., Isolauri, E., Salminen, S., Kumar, H., Satokari, R., and Virta, M. (2018). Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. Nature Communications *9,* 3891.

Partridge, S.R., Kwong, S.M., Firth, N., and Jensen, S.O. (2018). Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin. Microbiol. Rev. *31,* e00088-17.

Pei, R., Kim, S., Carlson, K.H., and Pruden, A. (2006). Effect of River Landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). Water Res. *40,* 2427-2435.

Phuong Hoa, P.T., Nonaka, L., Hung Viet, P., and Suzuki, S. (2008). Detection of the sul1, sul2, and sul3 genes in sulfonamide-resistant bacteria from wastewater and shrimp ponds of north Vietnam. Sci. Total Environ. *405,* 377-384.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments (FastTree 2). PLOS One *5,* e9490.

Qin, H., Wang, S., Feng, K., He, Z., Virta, M.P.J., Hou, W., Dong, H., and Deng, Y. (2019). Unraveling the diversity of sedimentary sulfate-reducing prokaryotes (SRP) across Tibetan saline lakes using epicPCR. Microbiome *7,* 71.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41,* D590.

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rissanen, A., Ojala, A., Fred, T., Toivonen, J., and Tiirola, M. (2017). Methylophilaceae and Hyphomicrobium as target taxonomic groups in monitoring the function of methanol-fed denitrification biofilters in municipal wastewater treatment plants. J. Ind. Microbiol. Biotechnol. *44,* 35-47.

Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M.C., Michael, I., and Fatta-Kassinos, D. (2013). Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. Sci. Total Environ. *447,* 345-360.

Roberts, M.C. (2008). Update on macrolide–lincosamide–streptogramin, ketolide, and oxazolidinone resistance genes. FEMS Microbiol. Lett. *282,* 147-159.

Roberts, M.C. (2005). Update on acquired tetracycline resistance genes. FEMS Microbiol. Lett. *245,* 195-203.

Roberts, M.C., Sutcliffe, J., Courvalin, P., Jensen, L.B., Rood, J., and Seppala, H. (1999). Nomenclature for Macrolide and Macrolide-Lincosamide-Streptogramin B Resistance Determinants. Antimicrob. Agents Chemother. *43,* 2823-2830.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4,* e2584.

Sandegren, L. (2014). Selection of antibiotic resistance at very low antibiotic concentrations. Upsala Journal of Medical Sciences *119,* 103-107.

Shen, Y., Stedtfeld, R.D., Guo, X., Bhalsod, G.D., Jeon, S., Tiedje, J.M., Li, H., and Zhang, W. (2019). Pharmaceutical exposure changed antibiotic resistance genes and bacterial communities in soil-surface- and overhead-irrigated greenhouse lettuce. Environ. Int. *131,* 105031.

Showsh, S.A., and Andrews, R.E. (1992). Tetracycline enhances Tn916-mediated conjugal transfer. Plasmid *28,* 213-224.

Spencer, S.J., Tamminen, M., Preheim, S., Guo, M., Briggs, A., Brito, I., Weitz, D., Pitkänen, L., Vigneault, F., Virta, M., and Alm, E. (2016). Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. The ISME Journal *10,* 427-436.

Sundin, G.W., and Bender, C.L. (1996). Dissemination of the strA-strB streptomycin-resistance genes among commensal and pathogenic bacteria from humans, animals, and plants. Molecular Ecology *5,* 133-143.

Tamminen, M., Karkman, A., Lõhmus, A., Muziasari, W.I., Takasu, H., Wada, S., Suzuki, S., and Virta, M. (2011). Tetracycline resistance genes persist at aquaculture farms in the absence of selection pressure. Environ. Sci. Technol. *45,* 386-391.

Thomas, C.M., and Nielsen, K.M. (2005). Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. Nature Reviews Microbiology *3,* 711-721.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, ISBN 978-3-319-24277-4.

Wu, L., Keucken, A., Wen, X., and Zhou, J. (2019). Global diversity and biogeography of bacterial communities in wastewater treatment plants. Nature Reviews Microbiology *4,* 1183-1195.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., and Larsen, M.V. (2012). Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. *67*, 2640-2644.

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614-620.

Zhao, W., and Hu, Z. (2013). Epidemiology and genetics of CTX-M extended-spectrum β-lactamases in Gram-negative bacteria. Critical Reviews in Microbiology *39*, 79-101.

Zhao, W., and Hu, Z. (2011). IMP-type metallo-β-lactamases in Gram-negative bacilli: distribution, phylogeny, and association with integrons. Critical Reviews in Microbiology *37*, 214.

APPENDIX

APPENDIX 1

Table: Primers used in this study

Name	Sequence	Usage	Source
Illum_341F_1	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**CCTACGGGNGGCWGCAG	16S rRNA gene forward primer with **Illumina TruSeq adapter**	Herlemann et al., 2011
Illum_341F_2	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**gtCCTACGGGNGGCWGCAG	16S rRNA gene forward primer with **Illumina TruSeq adapter**	Herlemann et al., 2011
Illum_341F_3	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**agagCCTACGGGNGGCWGCAG	16S rRNA gene forward primer with **Illumina TruSeq adapter**	Herlemann et al., 2011
Illum_341F_4	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**tagtgtCCTACGGGNGGCWGCAG	16S rRNA gene forward primer with **Illumina TruSeq adapter**	Herlemann et al., 2011
Illum_785R_1	**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**GACTACHVGGGTATCTAATCC	16S rRNA gene reverse primer; epicPCR nested reverse primer with **Illumina TruSeq adapter**	Herlemann et al., 2011
Illum_785R_2	**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**aGACTACHVGGGTATCTAATCC	16S rRNA gene reverse primer; epicPCR nested reverse primer with **Illumina TruSeq adapter**	Herlemann et al., 2011
Illum_785R_3	**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**tctGACTACHVGGGTATCTAATCC	16S rRNA gene reverse primer; epicPCR nested reverse primer with **Illumina TruSeq adapter**	Herlemann et al., 2011
Illum_785R_4	**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**ctgagtgGACTACHVGGGTATCTAATCC	16S rRNA gene reverse primer; epicPCR nested reverse primer with **Illumina TruSeq adapter**	Herlemann et al., 2011
pH'	AAGGAGGTGATCCAGCCGCA	epicPCR 16S rRNA gene reverse primer	Edwards et al., 1989
U519_blockF	TTTTTTTCAGCMGCCGCGGTAATWC/3SpC3/	epicPCR forward blocking primer	Spencer et al., 2016
U519_blockR	TTTTTTTGWATTACCGCGGCKGCTG/3SpC3/	epicPCR reverse blocking primer	Spencer et al., 2016
ermB-F1	TAAAGGGCATTTAACGACGAAACT	ermB epicPCR forward primer	Karkman et al., 2016
ermB-R1-F2'	**GWATTACCGCGGCKGCTG**CAGTTGACGATATTCTCGATTG	ermB epicPCR linker primer with **16S forward primer**	Karkman et al., 2016
ermB_F3_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**GGTTGCTCTTGCACACTCAAG	ermB epicPCR nested primer with **Illumina TruSeq adapter**	Karkman et al., 2016
ermF_F1	CAGCTTTGGTTGAACATTTACGAA	ermF epicPCR forward primer	Karkman et al., 2016
ermF_R1-F2'	**GWATTACCGCGGCKGCTG**AAATTCCTAAAATCACAACCGACAA	ermF epicPCR linker primer with **16S forward primer**	Karkman et al., 2016
ermF-F1_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**CAGCTTTGGTTGAACATTTACGAA	ermF epicPCR forward primer with **Illumina TruSeq adapter**	Karkman et al., 2016
blaIMP_AY242_F1	AAGGCAGCATTTCCTCTCATTTT	blaIMP epicPCR forward primer	Karkman et al., 2016
blaIMP_AY242_R1_F2´	**GWATTACCGCGGCKGCTG**GGATAGATCGAGAATTAAGCCACTCT	blaIMP epicPCR linker primer with **16S forward primer**	Karkman et al., 2016
blaIMP_AY242_F1_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**AAGGCAGCATTTCCTCTCATTTT	blaIMP epicPCR forward primer with **Illumina TruSeq adapter**	Karkman et al., 2016

ndm1_F1	CAACACAGCCTGACTTTCGC	blaNDM-1 epicPCR forward primer	University of Gothenburg (unpublished)
ndm1_R1_F2	**GWATTACCGCGGCKGCTG**TTGGCCTTGCTGTCCTTGATCA	blaNDM-1 epicPCR linker primer with **16S forward primer**	University of Gothenburg (unpublished)
ndm1_F3_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**ACACCAGTGACAATATCACCG	blaNDM-1 epicPCR nested primer with **Illumina TruSeq adapter**	University of Gothenburg (unpublished)
sul1_F1	GACGAGATTGTGCGGTTCTT	sul1 epicPCR forward primer	Department of Microbiology - University of Helsinki (unpublished)
sul1_R1_F2'	**GWATTACCGCGGCKGCTG**GAGACCAATAGCGGAAGCC	sul1 epicPCR linker primer with **16S forward primer**	Department of Microbiology - University of Helsinki (unpublished)
sul1_F3_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**ATGGGATTTTTCTTGAGCCC	sul1 epicPCR nested primer with **Illumina TruSeq adapter**	Department of Microbiology - University of Helsinki (unpublished)
strB_F1	CTAATGGCGAAGCTGTATG	strB epicPCR forward primer	Department of Microbiology - University of Helsinki (unpublished)
strB_R1_F2'	**GWATTACCGCGGCKGCTG**GTGGACGTAGTCAGTTTGAC	strB epicPCR linker primer with **16S forward primer**	Department of Microbiology - University of Helsinki (unpublished)
strB_F3_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**GTATGCCGCATCTGAGGAAC	strB epicPCR nested primer with **Illumina TruSeq adapter**	Department of Microbiology - University of Helsinki (unpublished)
tetM_F1	CATCATAGACACGCCAGGACA	tetM epicPCR forward primer	Karkman et al., 2016
tetM_R1_F2'	**GWATTACCGCGGCKGCTG**CTGTTTGATTACAATTTCCGC	tetM epicPCR linker primer with **16S forward primer**	Tamminen et al., 2011
tetM_F3_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**GCAATTCTACTGATTTCTGC	tetM epicPCR nested primer with **Illumina TruSeq adapter**	Tamminen et al., 2011
sul2_1_F1	CGGGAATGCCATCTGCCTTGAG	sul2 epicPCR forward primer, only tested	Pei et al., 2006
sul2_1_R1_F2	**GWATTACCGCGGCKGCTG**TCCGATGGAGGCCGGTATCTGG	sul2 epicPCR linker primer with **16S forward primer**, only tested	Karkman et al., 2016
sul2_1_F1_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**CGGGAATGCCATCTGCCTTGAG	sul2 epicPCR nested primer with **Illumina TruSeq adapter**, only tested	Pei et al., 2006
sul2_2_F1	TCCGATGGAGGCCGGTATCTGG	sul2 epicPCR forward primer, only tested	Karkman et al., 2016
sul2_2_R1_F2	**GWATTACCGCGGCKGCTG**CGGGAATGCCATCTGCCTTGAG	sul2 epicPCR linker primer with **16S forward primer**, only tested	Pei et al., 2006
sul2_2_F1_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**TCCGATGGAGGCCGGTATCTGG	sul2 epicPCR nested primer with **Illumina TruSeq adapter**, only tested	Karkman et al., 2016
sul2_3_F1	CGGGAATGCCATCTGCCTTGAG	sul2 epicPCR forward primer	Pei et al., 2006
sul2_3_R1_F2	**GWATTACCGCGGCKGCTG**TCCGATGGAGGCCGGTATCTGG	sul2 epicPCR linker primer with **16S forward primer**	Karkman et al., 2016
sul2_3_F3_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**cgatacgctcgatttctgtg	sul2 epicPCR nested primer with **Illumina TruSeq adapter**	Department of Microbiology - University of Helsinki (unpublished)
sul2_4_F1	TCCGATGGAGGCCGGTATCTGG	sul2 epicPCR forward primer, only tested	Karkman et al., 2016
sul2_4_R1_F2	**GWATTACCGCGGCKGCTG**CGGGAATGCCATCTGCCTTGAG	sul2 epicPCR linker primer with **16S forward primer**, only tested	Pei et al., 2006

sul2_4_F3_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**aggcgcgtaagctgatgg	sul2 epicPCR nested primer with **Illumina TruSeq adapter**, only tested	Department of Microbiology - University of Helsinki (unpublished)
tnpA-04_F1	CCGATCACGGAAAGCTCAAG	tnpA-04 epicPCR forward primer, only tested	Karkman et al., 2016
tnpA-04_R1-F2'	**GWATTACCGCGGCKGCTG**TGCTGCGAAATGGTGGTTG	tnpA-04 epicPCR linker primer with **16S forward primer**, only tested	Department of Microbiology - University of Helsinki (unpublished)
tnpA-04_F1_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**CCGATCACGGAAAGCTCAAG	tnpA-04 epicPCR forward primer with **Illumina TruSeq adapter**, only tested	Karkman et al., 2016
intI-1_3_F1	CGAAGTCGAGGCATTTCTGTC	intI-1 epicPCR forward primer, only tested	Karkman et al., 2016
intI-1_3_R1-F2´	**GWATTACCGCGGCKGCTG**GCCTTCCAGAAAACCGAGGA	intI-1 epicPCR linker primer with **16S forward primer**, only tested	Karkman et al., 2016
intI-1_3_F3_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**tgttcttctacggcaaggtg	intI-1 epicPCR nested primer with **Illumina TruSeq adapter**, only tested	Department of Microbiology - University of Helsinki (unpublished)
intI-1_4_F1	GCCTTCCAGAAAACCGAGGA	intI-1 epicPCR forward primer, only tested	Karkman et al., 2016
intI-1_4_R1-F2´	**GWATTACCGCGGCKGCTG**CGAAGTCGAGGCATTTCTGTC	intI-1 epicPCR linker primer with **16S forward primer**, only tested	Karkman et al., 2016
intI-1_4_F3_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**accttgccgtagaagaacag	intI-1 epicPCR nested primer with **Illumina TruSeq adapter**, only tested	Department of Microbiology - University of Helsinki (unpublished)
blaCTX-M-04_F1	CTTGGCGTTGCGCTGAT	blaCTX-M-04 epicPCR forward primer, only tested	Karkman et al., 2016
blaCTX-M-04_R1-F2´	**GWATTACCGCGGCKGCTG**CGTTCATCGGCACGGTAGA	blaCTX-M-04 epicPCR linker primer with **16S forward primer**, only tested	Karkman et al., 2016
blaCTX-M-04_F1_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**CTTGGCGTTGCGCTGAT	blaCTX-M-04 epicPCR forward primer with **Illumina TruSeq adapter**, only tested	Karkman et al., 2016
tetG_F1	CATCAGCGCCGGTCTTATG	tetG epicPCR forward primer, only tested	Karkman et al., 2016
tetG_R1-F2´	**GWATTACCGCGGCKGCTG**CCCCATGTAGCCGAACCA	tetG epicPCR linker primer with **16S forward primer**, only tested	Karkman et al., 2016
tetG_F1_TS	**ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**CATCAGCGCCGGTCTTATG	tetG epicPCR forward primer with **Illumina TruSeq adapter**, only tested	Karkman et al., 2016

Illumina TrueSeq adapters and 16S rRNA gene primer overhangs, when present, are indicated with bold characters.

In primers 341F and 785R, lower case letters indicate nucleotides that were added for mixing in sequencing.

APPENDIX 2

Table: Number of 16S rRNA gene reads obtained for each epicPCR sample and replicates that were chosen for analysis

Sample	Number of reads
blaNDM_i1	38680
blaNDM_i2	24904
blaNDM_i3	64
blaNDM_iN	2
blaNDM_e11	3671
blaNDM_e12	3
blaNDM_e21	820
blaNDM_e22	1865
blaNDM_e31	10208
blaNDM_e32	0
blaNDM_eN1	1
blaNDM_eN2	0
blaNDM_s11	520
blaNDM_s12	210
blaNDM_s13	0
blaNDM_s21	0
blaNDM_s22	48
blaNDM_s23	2
blaNDM_s31	358
blaNDM_sN1	0
blaNDM_sN2	8232
blaNDM_sN3	0

Sample	Number of reads
ermB_i1	23029
ermB_i2	19250
ermB_i3	7764
ermB_iN	10
ermB_e11	23197
ermB_e12	4847
ermB_e21	13032
ermB_e22	11
ermB_e31	8
ermB_e32	5
ermB_e33	68860
ermB_eN1	38
ermB_eN2	1
ermB_e3N	4
ermB_s11	865
ermB_s12	1848
ermB_s21	1272
ermB_s22	5306
ermB_s31	9
ermB_s32	1129
ermB_sN1	10
ermB_sN2	4

Sample	Number of reads
sul1_i1	122521
sul1_i2	138429
sul1_i3	115923
sul1_iN	48212
sul1_e11	80280
sul1_e12	62047
sul1_e21	1858
sul1_e22	4226
sul1_e31	927
sul1_e32	4593
sul1_eN1	246
sul1_eN2	10
sul1_s11	6490
sul1_s21	21
sul1_s22	173
sul1_s23	9
sul1_s31	22
sul1_s32	22
sul1_s33	7
sul1_sN1	15
sul1_sN2	22
sul1_sN3	2

Sample	Number of reads
blaIMP_i11	97365
blaIMP_i12	79228
blaIMP_i21	8887
blaIMP_i22	5427
blaIMP_i31	187073
blaIMP_i32	21770
blaIMP_iN1	7721
blaIMP_iN2	299
blaIMP_e11	7462
blaIMP_e12	2956
blaIMP_e21	2158
blaIMP_e22	6966
blaIMP_e31	10277
blaIMP_e32	10252
blaIMP_eN1	436
blaIMP_eN2	10
blaIMP_s11	23
blaIMP_s12	4764
blaIMP_s21	3
blaIMP_s22	1177
blaIMP_s31	74
blaIMP_s32	710
blaIMP_sN1	89
blaIMP_sN2	2490

Sample	Number of reads
ermF_i1	182135
ermF_i2	440895
ermF_i3	83793
ermF_iN	24
ermF_e11	5478
ermF_e12	29673
ermF_e21	351122
ermF_e22	76671
ermF_e31	77443
ermF_e32	49187
ermF_eN1	19
ermF_eN2	4
ermF_s11	479836
ermF_s21	40965
ermF_s31	208
ermF_s32	47283
ermF_sN1	36
ermF_sN2	14

Sample	Number of reads
strB_i11	25
strB_i12	137679
strB_i21	62
strB_i22	106
strB_i31	1259
strB_i32	74150
strB_iN1	0
strB_iN2	14
strB_e11	11
strB_e12	19
strB_e21	3
strB_e22	4
strB_e31	7
strB_e32	4
strB_eN1	1
strB_eN2	5
strB_s1	8
strB_s2	2
strB_s3	1
strB_sN	2

The samples names should to be read as: gene_ + letter indicating the sample type (i= influent, e = effluent, s = dried sludge) + number indicating the sampling day (biological replicate) or N for negative controls + number indicating the epicPCR technical replicate (absent if only one replicate was done). Highlighted cells represent samples that were considered in the analysis.